Remarks

The Amendments

Claim 1 has been amended to recite that a cluster map is generated by categorizing genes according to "unchanged, increased, or decreased expression levels" in place of "similarity in unchanged, increased, or decreased expression." The amendment is supported by the specification which discloses, "In some embodiments, cluster analysis methods are used to **group genes whose expression level is correlated**." (Page 40, lines 5-6, emphasis added.) The amendment is also supported by the specification which discloses an example of categorizing genes based on expression level. The specification discloses, "Comparison of hybridization signal intensities, that ranged over 4-orders of magnitude, revealed all categories of message expression changes including repressed (>10-fold down), down-regulated (<10-fold down), up-regulated (<10-fold up) and induced (>10-fold up) mRNAs between normal and malignant cells as shown in Fig. 10. (Page 52, line 33 to page 53, line 4.) Thus, the amendment is supported by the specification, introduces no new matter, and does not require further search or consideration. The amendment also does not narrow the scope of the claim. The amendment was not made earlier because claim 1 was first rejected as indefinite for reciting the amended phrase in the final Office Action. The amendment is also believed to place that claims in condition for allowance or better condition for appeal.

The Rejection of Claims 1-9, 106, and 107 Under 35 U.S.C. § 112, First Paragraph

Claims 1-9, 106, and 107 are rejected under 35 U.S.C. § 112, first paragraph, as not enabled by the specification. Applicants respectfully traverse.

Claim 1 is the only independent claim of the rejected claim set. Claim 1 recites a method for mapping a gene network. A plurality of biological samples is prepared. Expression of a least five genes is detected in the biological samples. Expression of the at least five genes in a first of said biological samples is compared to expression of the at least five genes in a second of said biological samples. A cluster map is generated for the genes. The cluster map categorizes genes according to unchanged, increased, or decreased expression levels in the first relative to the second of the biological samples. The cluster map is analyzed to generate gene network causal models defining regulatory relationships among the genes.

To satisfy the enablement requirement, the specification of a patent must teach those skilled in the art how to make and use the full scope of the claimed invention without undue experimentation. *In re Wright*, 999 F.2d 1557 (Fed. Cir. 1993). A specification need not disclose what is well known to those skilled in the art and preferably omits that which is well known to those skilled and already available to the public. *In re Buchner*, 929 F.2d 660, 661 (Fed. Cir. 1991); *Hybritech, Inc. v. Monoclonal Antibodies, Inc.*, 802 F.2d 1367, 1384 (Fed. Cir. 1986). Where different arts are involved in the invention, the specification is enabling if it enables persons skill in each art to carry out the aspect of the invention applicable to their specialty. *In re Naquin*, 398 F.2d 863 (C.C.P.A. 1968). Whenever an enablement rejection is made, the Patent Office must explain why it doubts

the truth or accuracy of any statement in a supporting disclosure and to back up assertions of its own with acceptable evidence or reasoning which is inconsistent with the contested statement. *In re Marzocchi*, 169 U.S.P.Q. (BNA) 367, 370 (C.C.P.A. 1971).

The Office Action asserts that claim 1, and thus dependent claims 2-9, 106, and 107, are not enabled because the specification and the art do not disclose how to perform the step of "analyzing said cluster map to generate gene network causal models defining regulatory relationships among said genes." (Paper 15, page 3, lines 4-6.) The Patent Office supports this assertion citing Schena (*Proc. Natl. Acad. Sci.* (1996) *93*:10614-10619) and Scientific Software International, Inc. (LISREL Historical Background, ©1997-2000, pages 1-3; "SSI"). Schena is cited as teaching the use of nucleic acid arrays to monitor gene expression, listing genes that are differentially expressed, and grouping the genes by expression level. Schena itself does not provide any teaching which indicates that one of skill in the art would be unable to perform the step of analyzing a cluster map. Scientific Software International, Inc. (SSI) is cited as teaching that "linear structural analysis includes algebraic formulation of the model in addition to the path diagram representation." (Paper 15, page 3, lines 20-21, quoting SSI at page 1, lines 33-34.) The Office Action reasons that the specification does not enable the step of analyzing because it does not provide "guidance (i.e. rules), algebraic algorithm(s), and path diagram representation for one of skill in the art to analyze the gene expression information provided by Schena et al. and generate a gene network causal model (i.e. using LISREL) thereby defining regulatory relationships among genes." (Paper 15, page 3, line 21 to page 47, line 2.) The Patent Office incorrectly assumes that one of ordinary skill in the art

would not have been able to analyze the cluster map without specific guidance (*i.e.*, rules), algebraic algorithm(s), or path diagram representations. SSI does not teach or suggest that the algebraic algorithm(s) or path analysis diagram representations needed to analyze any data are outside the knowledge of one of skill in the art. The art of statistics was well developed at the time the application was filed. One of skill in the art of statistics would have been readily able to analyze a cluster map to generate a gene network of causal models using the specification as a guide and the knowledge of how to perform analysis methods.

The specification discloses that Linear Structural Relations (LISREL) is one statistical method of analyzing clusters. "In some preferred embodiments, such cluster maps are then analyzed using statistical method to generate a map consisting of regulatory pathways describing the complex relationships among the genes. Many statistical methods are suitable for building such maps. The LISREL method is particularly useful in such application[s]." (Page 3, lines 4-9.) The specification also teaches how to perform the LISREL method. The specification discloses,

> Mathematical theories and applications of LISREL are described in detail in Joreskog and Sorbom (1979) Advances in Factor Analysis and Structural Equations Modeling, Cambridge MA, Abt Books; and Joreskog and Sorbom, (1985) LISREL IV: Analysis of Linear Structural Relationships by Maximum Likelihood Instrumental Variables and Least Squares, Uppsala, Sweden: University of Uppsala, **incorporated herein by reference for all purposes**.

Page 40, lines 27-32, emphasis added.

Advances in Factor Analysis and Structural Equations Modeling (Joreskog and Sorbom (1979); referred to hereafter as "Joreskog"; Exhibit A) teaches a general approach for performing LISREL and methods of testing if the causal model generated by LISREL exhibits goodness-of-fit, *i.e.*, if the observed data fit the causal model. Joreskog also teaches derivations of the general approach for performing LISREL and provides examples in which LISREL is used to generate models with empirical data.

Joreskog teaches that the general approach for performing LISREL is as follows:

> The models to be considered may all be formulated, in one way or another, in the following fashion. There are two sets of observed variables $y=(y_1, y_2, \ldots, y_p)'$ and $x=(x_1, x_2, \ldots, x_p)'$, here taken to be measured as deviations from their means. In some longitudinal models these are the post- and pre-tests, respectively. In other cases it is more convenient to regard all the observed variables as belonging to y, say, in which case the x-set is empty and we shall say that there is no x. It is assumed that the sets y and x satisfy factor analysis models with common factors $\eta=(\eta_1, \eta_2, \ldots, \eta_m)'$ and $\xi=(\xi_1, \xi_2, \ldots, \xi_n)'$ and unique factors $\varepsilon=(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_p)'$ and $\delta=(\delta_1, \delta_2, \ldots, \delta_q)'$ respectively, so that
>
> $$y=\Lambda_y\eta + \varepsilon, \qquad (1)$$
>
> and
>
> $$x=\Lambda_x\xi + \delta, \qquad (2)$$
>
> where Λ_y and Λ_x are factor loading matrices of order p x m and q x n, respectively. The usual assumptions of factor analysis are made, i.e.,
>
> $$E(\eta) = 0, E(\xi) = 0, E(\varepsilon) = 0, E(\delta) = 0, E(\eta\varepsilon') = 0,$$
> $$E(\xi\delta') = 0, E(\varepsilon\varepsilon') = \Theta^2_\varepsilon, E(\delta\delta') = \Theta^2_\delta, \text{ and } E(\varepsilon\delta') = 0,$$
>
> where Θ_ε and Θ_δ are diagonal matrices. The factors η and ξ are in general correlated (oblique) not only within sets but also between sets. The distinctive feature of LISREL is that a complete system of causal relationships may be postulated to hold among the factors, where the η's are treated as the jointly dependent (endogenous) variables and the ξ's as the independent (exogenous) variables. This

system is specified as a set of linear structural equations of the form

$$B\eta = \Gamma\xi + \zeta, \qquad (3)$$

where B and Γ are parameter matrices and ζ is a vector of random residuals (disturbance terms or errors in equations) assumed to be uncorrelated with ξ. In some models, B = I and (3) may be regarded as the multivariate regression of η on ξ. In other models when there is no x there is no ξ either, and (3) reduces to

$$B\eta = \zeta. \qquad (4)$$

Let Φ be the covariance matrix of ξ, and Ψ be the covariance matrix of ζ. Then the covariance matrix of the latent variables (η', ξ'), namely

$$\Omega = \begin{bmatrix} \Omega_{\eta\eta} & \Omega_{\eta\xi} \\ \Omega_{\xi\eta} & \Omega_{\xi\xi} \end{bmatrix},$$

Is a function of the elements of B, Γ, Φ, and Ψ:

$$\Omega_{\eta\eta} = B^{-1}\Gamma\Phi\Gamma' B'^{-1} + B^{-1}\Psi B'^{-1}, \qquad (5)$$
$$\Omega_{\eta\xi} = \Omega'_{\xi\eta} = B^{-1}\Gamma\Phi, \qquad (6)$$
$$\Omega_{\xi\xi} = \Phi. \qquad (7)$$

When there is no x, the covariance matrix Ω of η reduces to

$$\Omega = B^{-1}\Psi B'^{-1}. \qquad (8)$$

In the general case, the covariance matrix of the observed variables $(y', x')'$ implied by the model is

$$\Sigma = \begin{bmatrix} \Sigma_{yy} & \Sigma_{yx} \\ \Sigma_{xy} & \Sigma_{xx} \end{bmatrix},$$

Where

$$\Sigma_{yy} = \Lambda_y\Omega_{\eta\eta}\Lambda'_y + \Theta^2_\varepsilon \qquad (9)$$
$$\Sigma_{yx} = \Sigma'_{xy} = \Lambda_y\Omega_{\eta\xi}\Lambda'_x \qquad (10)$$
$$\Sigma_{xx} = \Lambda_x\Omega_{\xi\xi}\Lambda'_x\Theta^2_\delta. \qquad (11)$$

A model of the form of (1), (2), and (3), with the accompanying assumptions, has been called a LISREL model [Joreskog and van Thillo (1973)]. The LISREL model consists of two parts, the *measurement model* represented by (1) and (2) and the *structural equation model* represented by (3). The measurement model

specifies how the latent variables or hypothetical constructs are measured in terms of the observed variables. Its parameters Λ_x, Λ_y, Θ_δ, Θ_ε are used to describe the measurement properties (validities and reliabilities) of the observed variables. The structural equation model specified the causal relationships among the latent variables. Its parameters B, Γ and Ψ describe the causal effects and the amount of unexplained variance. A particular model is defined by specifying the nature of the elements of each of the parameter matrices $\Lambda_y(p \text{ x } m)$, $\Lambda_x(q \text{ x } n)$, $B(m \text{ x } m)$, $\Gamma(m \text{ x } n)$, $\Phi(n \text{ x } n)$, $\Psi(m \text{ x } m)$, $\Theta_\varepsilon(p \text{ x } p)$, and $\Theta_\delta(q \text{ x } q)$.

Page 130, line 28 to page 132, line 21. Joreskog teaches the basic equations that are required to perform a LISREL analysis. Joreskog also teaches how observed data is interrogated using the equations. Thus, Joreskog teaches how to generally perform the LISREL method. One of skill in the art would have been able to analyze a cluster map to generate gene network causal model using this basic LISREL method.

Furthermore, Joreskog teaches how to test if the observed data fits the model obtained using LISREL, *i.e.*, goodness-of-fit. Joreskog teaches a general equation that can be used to evaluate the goodness-of-fit of the model and that there is a computer program, LISREL II, that can estimate LISREL goodness-of-fit.

> Since we have assumed that the distribution of the observed variables is described by the covariance matrix, the estimation problem is essentially that of fitting the Σ imposed by the model to the sample covariance matrix S. As a fitting function we use
>
> $$F = \log|\Sigma| + \mathrm{tr}(S\Sigma^{-1}) - \log|S| - (p + q), \quad (13)$$
>
> which is to be minimized with respect to Θ. If the distribution of (y′, x′)′ is multinormal this yield maximum-likelihood estimates which are efficient in large samples. A computer program LISREL-II [Joreskog and van Thillo

(1973)] is available for computing the estimates of the LISREL parameters and their standard errors.

When the maximum-likelihood estimates have been obtained the computer program calculates the information matrix at the estimates. If this matrix is not positive definite the specified model is almost certainly not identified. (This is one way in which the identification problem can be handled if it has not been solved by other means.) For an identified model the information matrix will almost certainly be positive definite and can be inverted to give the standard errors.

The program also gives a χ^2-measure of overall goodness-of-fit of the model, which may be regarded as a test of the specified model against the most general alternative that Σ is an unconstrained positive definite matrix. The degrees of freedom for this χ^2-measure are

$$\tfrac{1}{2}(p + q)(p + q + 1) - s, \qquad (14)$$

where s is the number of identified parameters in the model.

Page 134, lines 4-26. Thus Joreskog teaches an evaluation of the causal model generated by the LISREL method that tests the goodness-of-fit of the observed data to the model.

Finally, Joreskog also teaches several variations of the LISREL method that can be used to analyze data. These variations include LISREL two-wave models (page 135-145), LISREL multi-wave, one- and two-variable models (page 145-159), and LISREL multi-wave, multi-variable models (page 159-167). Detailed descriptions for each method are provided as are methods for evaluating the goodness-of-fit of the models developed using the methods. Joreskog also teaches examples of the application of these models to data using the LISREL method for analysis. (Pages 156-159 and pages 161-166.) One of skill in the art, using these teachings, would have been able to "analyze said cluster map to generate gene network causal models defining regulatory relationships among said genes" and perform the method of claim 1 without recourse to undue experimentation.

The specification and the knowledge in the art at the time the application was filed were such that the skilled artisan would have been able to analyze a cluster map to generate gene network causal models defining regulatory relationships among genes. The specification discloses, by its incorporation of Joreskog, the analysis of a cluster map using LISREL. Joreskog clearly demonstrates that one of skill in the art of statistics would have been able to analyze a cluster map without recourse to undue experimentation.

Applicants respectfully request withdrawal of this rejection.

The Rejection of Claims 1-9, 106, and 107 Under 35 U.S.C. § 112, Second Paragraph

Claims 1-9, 106, and 107 have been rejected under 35 U.S.C. § 112, second paragraph, as not being definite.

The Office Action asserts that claim 1, and each of dependent claims 2-9, 106, and 107 are indefinite for reciting "the cluster map categorizes genes according to **similarity** in unchanged, increased, or decreased expression." (Emphasis added.) The Office Action asserts that the recitation is unclear because the range or degree of similarity in gene expression that is considered unchanged, increased, or decreased is unknown. (Paper 15, page 5, lines 2-3.) To advance prosecution, claim 1 has been amended to delete "similarity" from this recitation.

Applicants respectfully request withdrawal of the rejection.

Respectfully submitted,

Dated: October 16, 2003

By: [signature]
Michelle L. Holmes-Son
Registration No. 47,660

BANNER & WITCOFF, LTD.
1001 G Street, N.W., Eleventh Floor
Washington, D.C. 20001-4597
(202)824-3000

Advances in Factor Analysis and Structural Equation Models

Karl G. Jöreskog
Dag Sörbom

University of Uppsala
Uppsala, Sweden

with an Introduction by William W. Cooley

Edited by Jay Magidson

Abt Books
Cambridge, Massachusetts

Library of Congress Catalog Card Number 79-52433

Printed in the United States of America.

ISBN: 0-89011-535-4

Contents

Chapter 5

Statistical Models and Methods for Analysis of Longitudinal Data

Karl G. Jöreskog

1. Introduction

In the psychometric literature various models for the analysis of longitudinal data have been proposed and discussed. For example, Jöreskog (1970a), Werts, Jöreskog and Linn (1972), Corballis (1973), Schmidt and Wiley (1974), and Frederiksen (1974) have considered complex models which involve multiple measurements at several occasions. In the recent sociological literature a number of articles treat the analysis of data from panel studies [e.g., Bohrnstedt (1969), Heise (1969, 1970), Wiley and Wiley (1970), Duncan (1969, 1972)]. In this chapter we consider some of these models and various others from the point of view of formulation and statistical specification, estimation and testing.

The general setup is that of a longitudinal study where the same or similar quantitative measurements have been obtained at two or more occasions. Our models and examples are relevant for psychological and educational studies involving latent traits as well as for social and socio-economic studies involving unobserved variables or hypothetical constructs. In both types of studies there are several features that need to be considered.

(i) Most measurements employed in the behavioral and social sciences contain sizeable errors of measurement and any adequate model for the measurement of change must take this fact into account.
(ii) It is often difficult to obtain repeated measurements using the same measuring instrument (test) because of retest effects. When the measurements at the different occasions are not in the same units, scaling devices may be used to obtain approximately equal units.
(iii) When there are several measurements employed at each occasion, the question arises as to what traits are actually measured at each

Research reported in this paper has been supported by the Swedish Council for Social Science Research under project 'Statistical Methods for Analysis of Longitudinal Data', project director Karl G. Jöreskog.

occasion and how these are related. Special care must be taken to obtain measurements that actually measure the relevant latent traits or hypothetical constructs.

(iv) It may not be possible, nor even desirable, to specify the model completely. There may be other models which are equally plausible and a technique for choosing the most reasonable one is required. Further, if there is sufficient evidence to reject a given model the technique should suggest which part of the model is causing the poor fit.

In presenting our models it is convenient to use a path diagram. In the path diagram observed variables are enclosed in squares whereas latent (unobserved) variables are enclosed in circles. Residuals (errors in equations) and errors of measurements are included in the diagram but are not enclosed. A one-way arrow pointing from one variable x to another variable y indicates a possible direct causal influence of x on y, whereas a curved two-way arrow between x and y indicates that x and y may correlate without any causal interpretation of this correlation being given. The coefficient associated with each arrow is also displayed in the path diagram. For one-way arrows such coefficients will be (partial) regression coefficients. For two-way arrows they will be covariances, or if all variables are standardized, they will be correlations. With these conventions it is possible to write down the model equations from the path diagram.

2. The LISREL approach

2.1. Specification

Before proceeding to discuss various models suitable for longitudinal data, we first consider the general structure of all the models and methods of this paper. The models to be considered may all be formulated, in one way or another, in the following fashion. There are two sets of observed variables $y = (y_1, y_2, \ldots, y_p)'$ and $\boldsymbol{x} = (x_1, x_2, \ldots, x_q)'$, here taken to be measured as deviations from their means. In some longitudinal models these are the post- and pre-tests, respectively. In other cases it is more convenient to regard all the observed variables as belonging to y, say, in which case the $\boldsymbol{x}$-set is empty and we shall say that there is no $\boldsymbol{x}$. It is assumed that the sets y and $\boldsymbol{x}$ satisfy factor analysis models with common factors $\boldsymbol{\eta} = (\eta_1, \eta_2, \ldots, \eta_m)'$ and $\boldsymbol{\xi} = (\xi_1, \xi_2, \ldots, \xi_n)'$ and unique factors $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_p)'$ and $\boldsymbol{\delta} = (\delta_1, \delta_2, \ldots, \delta_q)'$, respectively, so that

$$y = \boldsymbol{\Lambda}_y \boldsymbol{\eta} + \boldsymbol{\varepsilon}, \tag{1}$$

and

$$\boldsymbol{x} = \boldsymbol{\Lambda}_x \boldsymbol{\xi} + \boldsymbol{\delta}, \tag{2}$$

where $\boldsymbol{\Lambda}_y$ and $\boldsymbol{\Lambda}_x$ are factor loading matrices of order $p \times m$ and $q \times n$, respectively. The usual assumptions of factor analysis are made, i.e.,

$$\begin{aligned}
&E(\boldsymbol{\eta}) = \boldsymbol{0}, \quad E(\boldsymbol{\xi}) = \boldsymbol{0}, \quad E(\boldsymbol{\varepsilon}) = \boldsymbol{0}, \quad E(\boldsymbol{\delta}) = \boldsymbol{0},\\
&E(\boldsymbol{\eta}\boldsymbol{\varepsilon}') = \boldsymbol{0}, \quad E(\boldsymbol{\xi}\boldsymbol{\delta}') = \boldsymbol{0}, \quad E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \boldsymbol{\Theta}_\varepsilon^2, \quad E(\boldsymbol{\delta}\boldsymbol{\delta}') = \boldsymbol{\Theta}_\delta^2, \quad E(\boldsymbol{\varepsilon}\boldsymbol{\delta}') = \boldsymbol{0},
\end{aligned}$$

where $\boldsymbol{\Theta}_\varepsilon$ and $\boldsymbol{\Theta}_\delta$ are diagonal matrices. The factors $\boldsymbol{\eta}$ and $\boldsymbol{\xi}$ are in general correlated (oblique) not only within sets but also between sets. The distinctive feature of LISREL is that a complete system of causal relationships may be postulated to hold among the factors, where the η's are treated as the jointly dependent (endogenous) variables and the ξ's as the independent (exogenous) variables. This system is specified as a set of linear structural equations of the form

$$\boldsymbol{B}\boldsymbol{\eta} = \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta}, \tag{3}$$

where $\boldsymbol{B}$ and $\boldsymbol{\Gamma}$ are parameter matrices and $\boldsymbol{\zeta}$ is a vector of random residuals (disturbance terms or errors in equations) assumed to be uncorrelated with $\boldsymbol{\xi}$. In some models, $\boldsymbol{B} = \boldsymbol{I}$ and (3) may be regarded as the multivariate regression of $\boldsymbol{\eta}$ on $\boldsymbol{\xi}$. In other models when there is no $\boldsymbol{x}$, there is no $\boldsymbol{\xi}$ either, and (3) reduces to

$$\boldsymbol{B}\boldsymbol{\eta} = \boldsymbol{\zeta}. \tag{4}$$

Let $\boldsymbol{\Phi}$ be the covariance matrix of $\boldsymbol{\xi}$, and $\boldsymbol{\Psi}$ be the covariance matrix of $\boldsymbol{\zeta}$. Then the covariance matrix of the latent variables $(\boldsymbol{\eta}', \boldsymbol{\xi}')'$, namely

$$\boldsymbol{\Omega} = \begin{pmatrix} \boldsymbol{\Omega}_{\eta\eta} & \boldsymbol{\Omega}_{\eta\xi} \\ \boldsymbol{\Omega}_{\xi\eta} & \boldsymbol{\Omega}_{\xi\xi} \end{pmatrix},$$

is a function of the elements of $\boldsymbol{B}$, $\boldsymbol{\Gamma}$, $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$:

$$\boldsymbol{\Omega}_{\eta\eta} = \boldsymbol{B}^{-1}\boldsymbol{\Gamma}\boldsymbol{\Phi}\boldsymbol{\Gamma}'\boldsymbol{B}'^{-1} + \boldsymbol{B}^{-1}\boldsymbol{\Psi}\boldsymbol{B}'^{-1}, \tag{5}$$

$$\boldsymbol{\Omega}_{\eta\xi} = \boldsymbol{\Omega}_{\xi\eta}' = \boldsymbol{B}^{-1}\boldsymbol{\Gamma}\boldsymbol{\Phi}, \tag{6}$$

$$\boldsymbol{\Omega}_{\xi\xi} = \boldsymbol{\Phi}. \tag{7}$$

When there is no $\boldsymbol{x}$, the covariance matrix $\boldsymbol{\Omega}$ of $\boldsymbol{\eta}$ reduces to

$$\boldsymbol{\Omega} = \boldsymbol{B}^{-1}\boldsymbol{\Psi}\boldsymbol{B}'^{-1}. \tag{8}$$

In the general case, the covariance matrix of the observed variables

$(y', x')'$ implied by the model is

$$\Sigma = \begin{pmatrix} \Sigma_{yy} & \Sigma_{yx} \\ \Sigma_{xy} & \Sigma_{xx} \end{pmatrix},$$

where

$$\Sigma_{yy} = \Lambda_y \Omega_{\eta\eta} \Lambda'_y + \Theta^2_\varepsilon, \qquad (9)$$

$$\Sigma_{yx} = \Sigma'_{xy} = \Lambda_y \Omega_{\eta\xi} \Lambda'_x, \qquad (10)$$

$$\Sigma_{xx} = \Lambda_x \Omega_{\xi\xi} \Lambda'_x + \Theta^2_\delta. \qquad (11)$$

A model of the form of (1), (2) and (3), with the accompanying assumptions, has been called a LISREL model [Jöreskog and van Thillo (1973)]. The LISREL model consists of two parts, the *measurement model* represented by (1) and (2) and the *structural equation model* represented by (3). The measurement model specifies how the latent variables or hypothetical constructs are measured in terms of the observed variables. Its parameters Λ_x, Λ_y, Θ_δ, Θ_ε are used to describe the measurement properties (validities and reliabilities) of the observed variables. The structural equation model specifies the causal relationships among the latent variables. Its parameters B, Γ and Ψ describe the causal effects and the amount of unexplained variance. A particular model is defined by specifying the nature of the elements of each of the parameter matrices Λ_y $(p \times m)$, Λ_x $(q \times n)$, B $(m \times m)$, Γ $(m \times n)$, Φ $(n \times n)$, Ψ $(m \times m)$, Θ_ε $(p \times p)$, and Θ_δ $(q \times q)$. Such an element may be specified as a

(i) fixed parameter, i.e., specified to have a given value,
(ii) free parameter, i.e., unknown to be estimated from data, or
(iii) constrained parameter, i.e., unknown but specified to be equal to one or more other parameters.

In the above presentation of the LISREL model it has been assumed that $\text{cov}(\delta_i, \delta_j) = 0$, $i \neq j$, $\text{cov}(\varepsilon_i, \varepsilon_j) = 0$, $i \neq j$, and $\text{cov}(\delta_i, \varepsilon_j) = 0$ for all i, j. These assumptions are not critical, however, since one can incorporate the δ's and ε's as factors along with the ξ's and η's, respectively. For example, to allow for correlated δ's one simply writes (2) as

$$x = (\Lambda_x\, I_q)\begin{pmatrix} \xi \\ \delta \end{pmatrix} + 0,$$

where I_q is the identity matrix of order q. Then $\Theta_\delta = 0$ and

$$\Phi = \begin{pmatrix} \Phi_{\xi\xi} & 0 \\ 0 & \Phi_{\delta\delta} \end{pmatrix},$$

where $\Phi_{\xi\xi}$ is the covariance matrix of ξ, and $\Phi_{\delta\delta}$ is the covariance matrix

of δ. A similar trick can be used to let the ε's be correlated. To allow the ε's to correlate with the δ's one simply uses the no-x option and treats all observed variables as y variables. Examples of such devices will be given later on in the chapter.

It should be noted that there is no requirement that $m < p$ and $n < q$ as in traditional factor analysis. The only requirement is that Σ is non-singular and that all parameters are identified.

2.2. *Identification*

We assume that the distribution of the observed variables is sufficiently well described by the moments of first and second order, so that information contained in moments of higher order may be ignored. In particular, this will hold if the distribution is multivariate normal. The distribution of $(y', x')'$ is therefore generated by the parameters in Λ_y, Λ_x, B, Γ, Φ, Ψ, Θ_ε, Θ_δ. Let θ be a vector of all the independent free and constrained parameters (i.e., counting each distinct constrained parameter only once) and let s be the order of θ. The identification problem then is the problem of whether or not θ is uniquely determined by Σ. Every θ generates one and only one Σ but two or more θ's could possibly generate the same Σ. If within the model there is only one θ for every Σ then θ is identified and we say that the whole model is identified. If, on the other hand, there are several θ's generating the same Σ we say that all such θ's are equivalent. If a parameter in θ has the same value in all equivalent θ's we say that this parameter is identified. If a parameter is not identified it will not be possible to find a consistent estimator of it.

To examine the identification of a model consider (9)–(11) after substitution of (5)–(7),

$$\sigma_{ij} = f_{ij}(\theta). \qquad (12)$$

There are $\frac{1}{2}(p+q)(p+q+1)$ equations and s unknown elements in θ. Hence a necessary condition for identification of all parameters is

$$s \leq \tfrac{1}{2}(p+q)(p+q+1).$$

If a parameter θ can be determined uniquely from Σ, this parameter is identified, otherwise it is not. Often some parameters can be determined from Σ in different ways. This gives rise to overidentifying conditions on Σ which must hold if the model is true. The solution of (12) is often complicated and explicit solutions for all θ's seldom exist. For certain special types of LISREL models general rules for the identification problem have been given by Fisher (1966), Wiley (1973), Geraci (1974),

and Werts, Jöreskog and Linn (1973). For most of our models we will consider the identification problem on a case-by-case basis.

2.3. *Estimation and testing*

Since we have assumed that the distribution of the observed variables is described by the covariance matrix, the estimation problem is essentially that of fitting the $\boldsymbol{\Sigma}$ imposed by the model to the sample covariance matrix $\boldsymbol{S}$. As a fitting function we use

$$F = \log|\boldsymbol{\Sigma}| + \operatorname{tr}(\boldsymbol{S}\boldsymbol{\Sigma}^{-1}) - \log|\boldsymbol{S}| - (p+q), \tag{13}$$

which is to be minimized with respect to $\boldsymbol{\theta}$. If the distribution of $(y', x')'$ is multinormal this yields maximum-likelihood estimates which are efficient in large samples. A computer program LISREL-II [Jöreskog and van Thillo (1973)] is available for computing the estimates of the LISREL parameters and their standard errors.

When the maximum-likelihood estimates have been obtained the computer program calculates the information matrix at the estimates. If this matrix is not positive definite the specified model is almost certainly not identified. (This is one way in which the identification problem can be handled if it has not been solved by other means.) For an identified model the information matrix will almost certainly be positive definite and can be inverted to give the standard errors.

The program also gives a χ^2-measure of overall goodness-of-fit of the model, which may be regarded as a test of the specified model against the most general alternative that $\boldsymbol{\Sigma}$ is an unconstrained positive definite matrix. The degrees of freedom for this χ^2-measure are

$$\tfrac{1}{2}(p+q)(p+q+1) - s, \tag{14}$$

where s is the number of identified parameters in the model.

Suppose H_0 represents one model under given specifications of fixed, free, and constrained parameters. To test the model H_0 against any more general model H_1, estimate them separately and compare their χ^2. The difference in χ^2 is asymptotically a χ^2 with degrees of freedom equal to the corresponding difference in degrees of freedom. In many situations, it is possible to set up a sequence of hypotheses such that each one is a special case of the preceding and to test these hypotheses sequentially. We will illustrate this procedure by means of several examples.

In a more exploratory situation the χ^2-goodness-of-fit-values can be

used as follows. If a value of χ^2 is obtained which is large compared to the number of degrees of freedom, the fit may be examined by an inspection of the residuals, i.e., the discrepancies between the observed and the reproduced variances and covariances. That examination, in conjunction with subject-matter considerations, may suggest ways to relax the model somewhat by introducing more parameters. The new model usually yields a smaller χ^2. A large drop in χ^2, compared to the difference in degrees of freedom, supports the changes made. On the other hand, a drop in χ^2 which is close to the difference in number of degrees of freedom indicates that the improvement in fit is obtained by capitalizing on chance.

3. Two-wave models

3.1. *Two-wave, two-variable models*

Consider the situation where we have two observed variables at each of two occasions. Such a two-wave, two-variable model is shown in Figure 1a, where y_1 and y_2 are measured at the first occasion and y_3 and y_4 at the second occasion. To fix the scales for the latent variables η_1 and η_2 we assume that these are measured in the same metric as y_1 and y_3, respectively. The measurement model is

$$\begin{aligned} y_1 &= \eta_1 + \varepsilon_1, & y_2 &= \lambda_1\eta_1 + \varepsilon_2, \\ y_3 &= \eta_2 + \varepsilon_3, & y_4 &= \lambda_2\eta_2 + \varepsilon_4. \end{aligned} \tag{15}$$

The implied covariance matrix of $y = (y_1, y_2, y_3, y_4)'$ is

$$\boldsymbol{\Sigma} = \begin{bmatrix} \omega_{11} + \theta_{11}^{(\varepsilon)} & & & \\ \lambda_1\omega_{11} & \lambda_1^2\omega_{11} + \theta_{22}^{(\varepsilon)} & & \\ \omega_{21} & \lambda_1\omega_{21} & \omega_{22} + \theta_{33}^{(\varepsilon)} & \\ \lambda_2\omega_{21} & \lambda_1\lambda_2\omega_{21} & \lambda_2\omega_{22} & \lambda_2^2\omega_{22} + \theta_{44}^{(\varepsilon)} \end{bmatrix}, \tag{16}$$

where

$$\boldsymbol{\Omega} = \begin{pmatrix} \omega_{11} & \omega_{12} \\ \omega_{21} & \omega_{22} \end{pmatrix}$$

is the covariance matrix of η_1 and η_2. The matrix $\boldsymbol{\Sigma}$ has 10 distinct elements which are expressed in terms of 9 parameters. It is clear that $\omega_{21} = \sigma_{31}$ is identified. With ω_{21} determined, λ_1 is determined from σ_{32} and

λ_2 from σ_{41}. The constraint $\sigma_{42} = \lambda_1\lambda_2\omega_{21}$ represents one overidentifying restriction. The variances ω_{11} and ω_{22} are determined from σ_{21} and σ_{43}, respectively. Finally, $\theta_{ii}^{(\varepsilon)}$ are determined from σ_{ii} $(i = 1, \ldots, 4)$. Hence this model is identified with one overidentifying restriction.

When the same measuring instrument (test or questionnaire) is used at both occasions there is usually a tendency for the errors in each variable to correlate over time because of memory or other re-test effects. In this situation the model should allow for correlations between ε_1 and ε_3 and also between ε_2 and ε_4. Such a model is illustrated in Figure 1b. Duncan (1972) and Kenny (1973) have previously considered models of this type. The covariance matrix Σ for the model in Figure 1b is the same as (16) except that $\theta_{31}^{(\varepsilon)} = \text{cov}(\varepsilon_1, \varepsilon_3)$ and $\theta_{42}^{(\varepsilon)} = \text{cov}(\varepsilon_2, \varepsilon_4)$ are added to σ_{31} and σ_{42}, respectively. There are now 11 unknown parameters but only 10 observable moments. Hence the model is not identified.

Indeed none of the 11 parameters are identified without additional conditions. The loadings λ_1 and λ_2 may be multiplied by a constant and the ω's divided by the same constant, without changing σ_{21}, σ_{32}, σ_{41}, and σ_{43}. The changes in the other σ's may be compensated by adjusting the θ's additively. To make the model identified one must fix one λ or one ω at a non-zero value, or one θ at some arbitrary value. There does not seem to be any reasonable way to do so, unless one is willing to make some further assumptions about the nature of the variables. For example, if the variables y_1 and y_2 are tau-equivalent [see, e.g., Lord and Novick (1968, p.

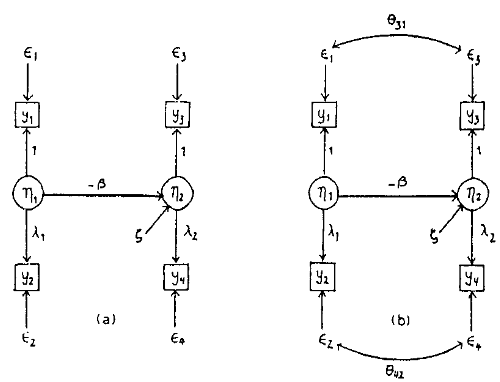

FIGURE 1. Two-wave, two-variable model with (a) uncorrelated and (b) correlated errors.

47)] we set $\lambda_1 = 1$. The model is then just-identified and will fit most data exactly. However, as pointed out by Duncan (1972) and Kenny (1973), the *correlation* between η_1 and η_2 is identified without any restriction, because

$$\rho^2_{\eta_1\eta_2} = \omega^2_{21}/\omega_{11}\omega_{22} = (\sigma_{41}\sigma_{32})/(\sigma_{21}\sigma_{43}).$$

The model may be used to estimate this correlation and to test whether it is unity. When the model is just-identified, the maximum-likelihood estimate of $\rho^2_{\eta_1\eta_2}$ is

$$\hat{\rho}^2_{\eta_1\eta_2} = (s_{41}s_{32})/(s_{21}s_{43}),$$

where s_{ij} is the sample covariance corresponding to σ_{ij}.

3.2. *Background variables*

In many longitudinal studies the objective is to measure change between two occasions and to relate the change to certain characteristics and events. Such studies include not only pre- and post-measures but also various background variables believed to influence change. The background variables are socio-economic or other characteristics which differentiate the individuals prior to the pre-test occasion.

Consider the model shown in Figure 2. The background variable, denoted x, is the only independent (exogenous) variable. The main purpose of the model is to estimate the direct effect of η_1 on η_2 by eliminating the effect of x. We shall consider this model under two alternative conditions on the background variable, namely,

(i) x is measured without error,
(ii) x is fallible.

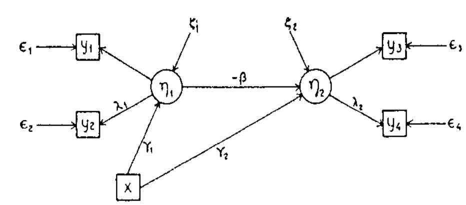

FIGURE 2. Two pre-tests, two post-tests, and an infallible background variable.

3.3. *Accurate background variable*

First suppose x is measured without error. The measurement model for $y' = (y_1, y_2, y_3, y_4)$ is

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} = \begin{bmatrix} 1 & 0 \\ \lambda_1 & 0 \\ 0 & 1 \\ 0 & \lambda_2 \end{bmatrix} \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \end{pmatrix}, \tag{17}$$

and the structural model for η_1, η_2, and x is

$$\begin{pmatrix} 1 & 0 \\ \beta & 1 \end{pmatrix} \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix} = \begin{pmatrix} \gamma_1 \\ \gamma_2 \end{pmatrix} x + \begin{pmatrix} \zeta_1 \\ \zeta_2 \end{pmatrix}. \tag{18}$$

Solving (18) for η_1 and η_2 we get

$$\eta_1 = \gamma_1 x + \zeta_1,$$
$$\eta_2 = (\gamma_2 - \beta\gamma_1)x + (\zeta_2 - \beta\zeta_1)$$
$$= \pi x + \upsilon,$$

say.

Let us first consider the identification problem. We have five observed variables y_1, y_2, y_3, y_4, and x, so that we have fifteen variances and covariances in Σ. We have twelve unknown parameters: λ_1, λ_2, β, γ_1, γ_2, $\phi = \text{var}(x)$, $\psi_{11} = \text{var}(\zeta_1)$, $\psi_{22} = \text{var}(\zeta_2)$, and $\theta^2_{\varepsilon_i} = \text{var}(\varepsilon_i)$, $i = 1, 2, 3, 4$. We have

$$\text{cov}(y_1, x) = \text{cov}(\eta_1, x) = \gamma_1\phi, \tag{19}$$
$$\text{cov}(y_2, x) = \lambda_1 \text{cov}(\eta_1, x) = \lambda_1\gamma_1\phi, \tag{20}$$
$$\text{cov}(y_3, x) = \text{cov}(\eta_2, x) = \pi\phi, \tag{21}$$
$$\text{cov}(y_4, x) = \lambda_2 \text{cov}(\eta_2, x) = \lambda_2\pi\phi. \tag{22}$$

Since $\phi = \text{var}(x)$ is identified, these equations determine γ_1, λ_1, π and λ_2, respectively. Furthermore,

$$\text{cov}(y_1, y_2) = \lambda_1 \text{var}(\eta_1) = \lambda_1(\gamma_1^2\phi + \psi_{11}) \tag{23}$$

determines ψ_{11}, and

$$\text{cov}(y_3, y_4) = \lambda_2 \text{var}(\eta_2) = \lambda_2[\pi^2\phi + \text{var}(\upsilon)] \tag{24}$$

determines

$$\text{var}(\upsilon) = \psi_{22} + \beta^2\psi_{11}. \tag{25}$$

With λ_1, λ_2, $\gamma_1\pi$, ϕ and ψ_{11} in hand, any one of the four equations

$$\text{cov}(y_1, y_3) = \gamma_1\pi\phi - \beta\psi_{11}, \tag{26}$$
$$\text{cov}(y_1, y_4) = \lambda_2(\gamma_1\pi\phi - \beta\psi_{11}), \tag{27}$$
$$\text{cov}(y_2, y_3) = \lambda_1(\gamma_1\pi\phi - \beta\psi_{11}), \tag{28}$$
$$\text{cov}(y_2, y_4) = \lambda_1\lambda_2(\gamma_1\pi\phi - \beta\psi_{11}), \tag{29}$$

can be used to determine β. Thus there are three overidentifying restrictions. Then, with β determined, $\gamma_2 = \pi + \beta\gamma_1$ and ψ_{22} is obtained from (25). Finally the error variances $\theta^2_{\varepsilon_i}$ are determined from $\text{var}(y_i)$, $i = 1, 2, 3, 4$. Hence the whole model is identified and there are three independent restrictions on Σ.

3.4. *Fallible background variable*

Now suppose x is fallible,

$$x = \xi + \delta,$$

where ξ is the true score and δ the measurement error, the latter assumed to have zero mean and to be uncorrelated with ξ and all other unobserved variables. We shall consider two cases, namely; (a) x has a known reliability,

$$\rho_{xx} = \text{var}(\xi)/\text{var}(x),$$

and (b) ξ is measured by two congeneric background variables x_1 and x_2.

The first case is shown in Figure 3 and is similar to the model just analyzed. In (19) through (29) everything is the same except that $\text{var}(x)$ is replaced by $\text{var}(\xi)$. Since $\text{var}(\xi) = \rho_{xx}\,\text{var}(x)$, where ρ_{xx} is known and $\text{var}(x)$ is identified, all the other parameters will be determined as before.

An extension of the model in Figure 3 to include correlations between ε_1 and ε_3 and between ε_2 and ε_4 is shown in Figure 4. Then $\text{cov}(\varepsilon_1, \varepsilon_3)$ will be added to the right side of (26) and $\text{cov}(\varepsilon_2, \varepsilon_4)$ will be added to the right side of (29). For this model, λ_1, λ_2, γ_1, π and ϕ are determined as before. Then (27) and (28) determine β with one overidentifying restriction, and $\text{cov}(\varepsilon_1, \varepsilon_3)$ and $\text{cov}(\varepsilon_2, \varepsilon_4)$ are then uniquely determined by (26) and (29), respectively. Hence this model is identified and there is one restriction on Σ.

The case of two congeneric background variables is shown in Figure 5.

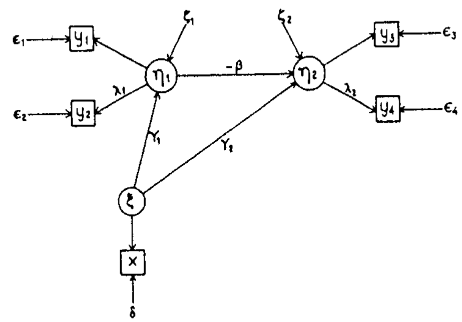

FIGURE 3. Two pre-tests, two post-tests, and a fallible background variable.

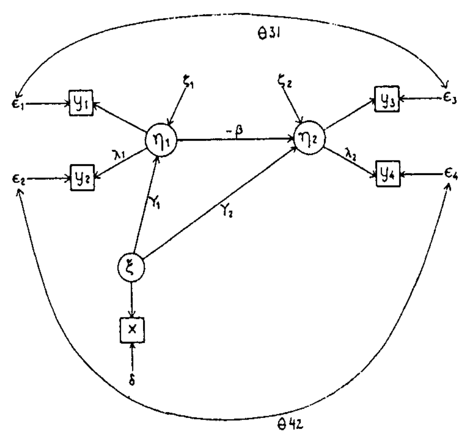

FIGURE 4. Two pre-tests, two post-tests, a fallible background variable, and correlated errors.

Here we write

$$x_1 = \xi + \delta_1, \tag{30}$$

$$x_2 = \lambda_3 \xi + \delta_2, \tag{31}$$

where λ_3 is a parameter to be determined, and δ_1 and δ_2 are uncorrelated measurement errors, uncorrelated with ξ and all other latent variables.

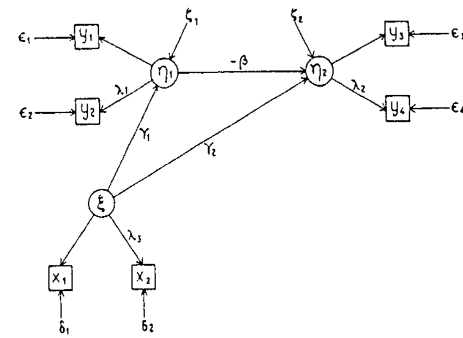

FIGURE 5. Two pre-tests, two post-tests, and two congeneric background variables.

The other equations are as before except that in (18) x is replaced by ξ. We now have three more parameters than before, namely, λ_3, $\theta^2_{\delta_1} = \text{var}(\delta_1)$, and $\theta^2_{\delta_2} = \text{var}(\delta_2)$. The new parameter $\phi = \text{var}(\xi)$ replaces the old $\phi = \text{var}(x)$. On the other hand we now have six more observable moments, so that the model has six degrees of freedom when $\text{cov}(\varepsilon_1, \varepsilon_3) = \text{cov}(\varepsilon_2, \varepsilon_4) = 0$, and four degrees of freedom when these covariances are left free.

The parameter λ_3 is identified with three overidentifying restrictions since

$$\text{cov}(x_2, y_i)/\text{cov}(x_1, y_i) = \lambda_3, \qquad i = 1, 2, 3, 4.$$

All the other parameters are determined as before.

3.5. *Estimation*

We discuss estimation of the model of Figure 5 explicitly; the other cases can be handled in a similar manner.

When the ε's are all uncorrelated, the model can be directly specified in the LISREL format. We have

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ \lambda_1 & 0 \\ 0 & 1 \\ 0 & \lambda_2 \end{pmatrix} \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \end{pmatrix}, \tag{32}$$

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 \\ \lambda_3 \end{pmatrix} \xi + \begin{pmatrix} \delta_1 \\ \delta_2 \end{pmatrix}, \tag{33}$$

$$\begin{pmatrix} 1 & 0 \\ \beta & 1 \end{pmatrix} \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix} = \begin{pmatrix} \gamma_1 \\ \gamma_2 \end{pmatrix} \xi + \begin{pmatrix} \zeta_1 \\ \zeta_2 \end{pmatrix}. \tag{34}$$

The LISREL program gives maximum-likelihood estimates of λ_1, λ_2, λ_3, β, γ_1, γ_2, ψ_{11}, ψ_{22}, ϕ, $\theta^2_{\delta_i}$ $(i = 1, 2)$, and $\theta^2_{\varepsilon_i}$ $(i = 1, 2, 3, 4)$ and standard errors for these estimates.

When cov $(\varepsilon_1, \varepsilon_3)$ and cov $(\varepsilon_2, \varepsilon_4)$ are unrestricted, we incorporate the ε's as ξ's and η's as follows:

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} = \begin{bmatrix} 1 & 0 & 1 & 0 & 0 & 0 \\ \lambda_1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & \lambda_2 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} \eta_1 \\ \eta_2 \\ \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \end{pmatrix}, \tag{35}$$

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ \lambda_3 & 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \xi \\ \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \end{pmatrix} + \begin{pmatrix} \delta_1 \\ \delta_2 \end{pmatrix}, \tag{36}$$

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ \beta & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} \eta_1 \\ \eta_2 \\ \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \end{pmatrix} = \begin{bmatrix} \gamma_1 & 0 & 0 & 0 & 0 \\ \gamma_2 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} \xi \\ \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \end{pmatrix} + \begin{pmatrix} \zeta_1 \\ \zeta_2 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}. \tag{37}$$

In (35) we represent the ε's as additional η's and in (36) we also represent them as additional ξ's.

So in LISREL we take

$$\Phi = \begin{pmatrix} \phi & & & & \\ 0 & \phi_{11} & & & \\ 0 & 0 & \phi_{22} & & \\ 0 & \phi_{31} & 0 & \phi_{33} & \\ 0 & 0 & \phi_{42} & 0 & \phi_{44} \end{pmatrix},$$

where $\phi = \text{var}(\xi)$, $\phi_{ii} = \text{var}(\varepsilon_i)$ and $\phi_{ij} = \text{cov}(\varepsilon_i, \varepsilon_j)$, $i, j = 1, 2, 3, 4$, and

$$\Psi = \begin{bmatrix} \psi_{11} & & & & & \\ 0 & \psi_{22} & & & & \\ 0 & 0 & 0 & & & \\ 0 & 0 & 0 & 0 & & \\ 0 & 0 & 0 & 0 & 0 & \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

An example of this model follows in the next section.

3.6. *The stability of alienation*

To illustrate the theory we draw on the data in Wheaton et al. (1975). This study was concerned with the stability over time of attitudes such as alienation and the relation of attitudes to background variables such as education and occupation. Data on attitude scales were collected from 932 persons in two rural regions in Illinois at three points in time: 1966, 1967 and 1971 [see Summers et al. (1969) for further description of the research setting]. We use only the data for 1967 and 1971. The variables we use are the *Anomia* subscale and the *Powerlessness* subscale, taken to be indicators of *Alienation*. The background variables are respondent's education (years of schooling completed) and Duncan's Socio-Economic Index (SEI). These are taken to be indicators of respondent's Socio-Economic Status (SES). We analyze the data under the three models shown in Figures 6a–c, none of which correspond to Wheaton's final model.

The maximum-likelihood estimates of the parameters are given in Table 1. The main aim of the Wheaton study was to estimate the stability of alienation over time, which is reflected in the parameter β, or rather in the squared correlation between ALIENATION 71 and ALIENATION 67.

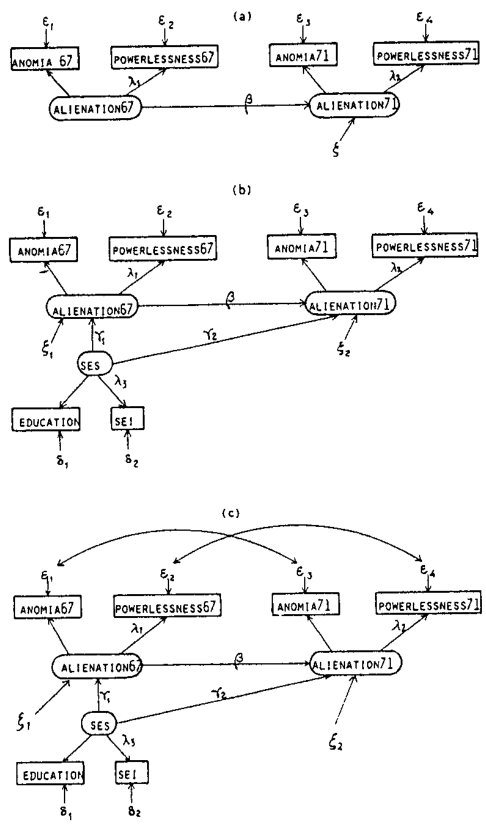

FIGURE 6. Models for study of stability of alienation.

The squared correlation coefficients are 0.56, 0.56 and 0.46 for Models a, b and c, respectively.

As can be seen from Table 1, β is overestimated in Model 6a, which ignores SES. Model 6b indicates that SES has a significant influence on alienation on both occasions. The negative signs of the SES coefficients γ_1 and γ_2 indicate that high socio-economic status reduces alienation. However, the overall fit of Model 6b is not acceptable. Since the same scales are used on both occasions, it seems reasonable to allow error correlations between the same measures at the two occasions. This is

TABLE 1

Parameter estimates for the models in Figure 6a–c (standard errors in parentheses).

Parameter	Model 6a	Model 6b	Model 6c
λ_1	0.815 (0.040)	0.888 (0.041)	0.979 (0.062)
λ_2	0.847 (0.042)	0.849 (0.040)	0.922 (0.059)
λ_3	-	5.331 (0.430)	5.221 (0.422)
β	0.789 (0.044)	0.705 (0.054)	0.607 (0.051)
γ_1	-	−0.614 (0.056)	−0.575 (0.056)
γ_2	-	−0.174 (0.054)	−0.227 (0.052)
ψ_{11}	-	5.307 (0.473)	4.847 (0.468)
ψ_{22}	4.085 (0.432)	3.742 (0.388)	4.089 (0.405)
ϕ	-	6.663 (0.641)	6.803 (0.650)
σ_{δ_1}	-	1.717 (0.145)	1.675 (0.151)
σ_{δ_2}	-	16.153 (0.565)	16.273 (0.558)
σ_{ϵ_1}	1.906 (0.097)	2.004 (0.086)	2.176 (0.104)
σ_{ϵ_2}	1.865 (0.077)	1.786 (0.076)	1.602 (0.126)
σ_{ϵ_3}	1.827 (0.109)	1.923 (0.097)	2.098 (0.123)
σ_{ϵ_4}	1.969 (0.077)	1.904 (0.077)	1.754 (0.124)
corr (ϵ_1, ϵ_3)	-	-	0.356 (0.047)
corr (ϵ_2, ϵ_4)	-	-	0.121 (0.082)
χ^2	61.155	71.544	4.770
d.f.	1	6	4

done in Model 6c, which, as can be seen in Table 1, provides an acceptable fit.

4. Multi-wave, one- and two-variable models

We now turn to the multi-wave case with one or two variables being measured on the subjects on three or more occasions.

4.1. *Multi-wave, one-variable models*

Suppose one fallible measure y is administered repeatedly to the same group of people. A possible model for this situation is shown in Figure 7. (We confine attention to the four-occasion case here, the extension to the general case being obvious throughout.) Such schemes have been termed simplex models by Guttman (1954) because of the typical pattern of intercorrelations they produce. Anderson (1960) formulated this model in terms of various stochastic processes and treated the identification problem; Jöreskog (1970b) treated the estimation problem. Applications

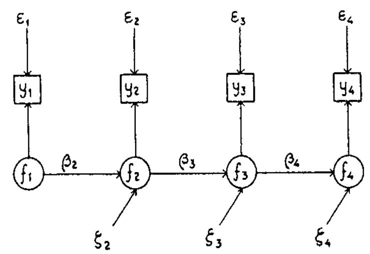

FIGURE 7. A simplex model.

to sociological panel analysis have been discussed by Heise (1969), Wiley and Wiley (1970), and Werts, Jöreskog and Linn (1971), and an application to the measurement of academic growth has been given by Werts, Linn and Jöreskog (1976).

The unit of measurement in each factor f_i is chosen to be the same as in the corresponding y_i. The equations defining the model are

$$y_i = f_i + \varepsilon_i, \qquad i = 1, 2, 3, 4, \tag{38}$$

$$f_{i+1} = \beta_{i+1} f_i + \zeta_{i+1}, \qquad i = 1, 2, 3. \tag{39}$$

Here the ε_i are mutually uncorrelated and uncorrelated with all the f_i, and ζ_{i+1} is uncorrelated with f_i. The unknown parameters of the model are $\phi_i = \text{var}(f_i)$, $\theta_i^2 = \text{var}(\varepsilon_i)$, $i = 1, 2, 3, 4$, and $\beta_2, \beta_3, \beta_4$. (We use the symbol ϕ rather than ω here to denote the variances of the dependent variables. Since there are no independent variables, there should be no confusion.) The residual variances are $\text{var}(\zeta_{i+1}) = \phi_{i+1} - \beta_{i+1}^2 \phi_i$, $i = 1, 2, 3$. The covariance matrix of y_1, y_2, y_3 and y_4 is

$$\boldsymbol{\Sigma} = \begin{bmatrix} \phi_1 + \theta_1^2 & & & \\ \beta_2\phi_1 & \phi_2 + \theta_2^2 & & \\ \beta_2\beta_3\phi_1 & \beta_3\phi_2 & \phi_3 + \theta_3^2 & \\ \beta_2\beta_3\beta_4\phi_1 & \beta_3\beta_4\phi_2 & \beta_4\phi_3 & \phi_4 + \theta_4^2 \end{bmatrix}. \tag{40}$$

It is seen from (40) that the product $\beta_2\phi_1 = \sigma_{21}$ is identified, but β_2 and ϕ_1 are not separately identified. We can multiply β_2 by a constant and divide ϕ_1 by the same constant without changing the product; the change induced in ϕ_1 can be absorbed in θ_1^2 in such a way that σ_{11} remains unchanged. Hence $\theta_1^2 = \text{var}(\varepsilon_1)$ is not identified either. On the other hand,

for f_2 and f_3 we have

$$\phi_2 = (\sigma_{32}\sigma_{21})/\sigma_{31}, \qquad \phi_3 = (\sigma_{43}\sigma_{32})/\sigma_{42},$$

so that ϕ_2 and ϕ_3, and hence also θ_2^2 and θ_3^2, are identified. With ϕ_2 and ϕ_3 in hand, β_3 and β_4 are determined by σ_{32} and σ_{43}. The central coefficient β_3 is overidentified since

$$\beta_3\phi_2 = (\sigma_{31}\sigma_{42})/\sigma_{41} = \sigma_{32}.$$

Since both ϕ_4 and θ_4^2 enter only into σ_{44}, only their sum is identified.

This analysis shows that for the 'inner' variables (y_2 and y_3), the parameters ϕ_2, ϕ_3, θ_2, θ_3 and β_3 are identified, whereas there is an indeterminacy associated with the 'outer' variables (y_1 and y_4). To eliminate these indeterminacies one of the parameters ϕ_1, θ_1 and β_2 must be specified and one of the parameters ϕ_4 and θ_4 must also be specified. Hence there are only nine independent parameters and the model has one degree of freedom. In the general case of $m \geq 4$ occasions there will be $3m - 3$ free parameters and the degrees of freedom is $\frac{1}{2}m(m+1) - (3m - 3)$.

The simplex model can be put into the LISREL format, with no $\boldsymbol{x}$. We write

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} f_1 \\ f_2 \\ f_3 \\ f_4 \end{bmatrix} + \begin{pmatrix} 0 \\ \varepsilon_2 \\ \varepsilon_3 \\ 0 \end{pmatrix}, \tag{41}$$

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ -\beta_2 & 1 & 0 & 0 \\ 0 & -\beta_3 & 1 & 0 \\ 0 & 0 & -\beta_4 & 1 \end{bmatrix} \begin{bmatrix} f_1 \\ f_2 \\ f_3 \\ f_4 \end{bmatrix} = \begin{bmatrix} \zeta_1 \\ \zeta_2 \\ \zeta_3 \\ \zeta_4 \end{bmatrix}. \tag{42}$$

In (41) we have taken $\varepsilon_1 = \varepsilon_4 = 0$ to eliminate the indeterminacies and in (42) we have defined ζ_1 as f_1. Note that we have taken $\psi_i = \text{var}(\zeta_i)$ as free parameters, rather than the $\phi_i = \text{var}(f_i)$; there is obviously a one-to-one correspondence between them. The parameter matrices are specified as

$\boldsymbol{\Lambda}_y = \boldsymbol{I}$,

$\boldsymbol{B}$ is as in (42),

$$\boldsymbol{\Psi} = \mathrm{diag}\,(\psi_1, \psi_2, \psi_3, \psi_4),$$

$$\boldsymbol{\Theta}_\epsilon = \mathrm{diag}\,(0, \sigma_{\epsilon_2}, \sigma_{\epsilon_3}, 0).$$

4.2. *Multi-wave, two-variable models*

Now suppose that we have two observed variables at each of four occasions as in Figure 8. This model is a direct generalization of the model in Figure 1b to four occasions.

With $\boldsymbol{x}' = (x_1, x_2, x_3, x_4)$, $\boldsymbol{y}' = (y_1, y_2, y_3, y_4)$, the measurement model is

$$\boldsymbol{x} = \boldsymbol{f} + \boldsymbol{\delta}, \tag{43}$$

$$\boldsymbol{y} = \boldsymbol{D}_\lambda \boldsymbol{f} + \boldsymbol{\varepsilon}, \tag{44}$$

where $\boldsymbol{D}_\lambda = \mathrm{diag}\,(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$. The covariance matrix of $\boldsymbol{z} = (\boldsymbol{x}', \boldsymbol{y}')'$ is

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{xx} & \boldsymbol{\Sigma}_{xy} \\ \boldsymbol{\Sigma}_{yx} & \boldsymbol{\Sigma}_{yy} \end{pmatrix}, \tag{45}$$

with

$$\boldsymbol{\Sigma}_{xx} = \boldsymbol{\Phi} + \boldsymbol{\Theta}^*_\delta, \tag{46}$$

$$\boldsymbol{\Sigma}_{yx} = \boldsymbol{D}_\lambda \boldsymbol{\Phi}, \tag{47}$$

$$\boldsymbol{\Sigma}_{yy} = \boldsymbol{D}_\lambda \boldsymbol{\Phi} \boldsymbol{D}_\lambda + \boldsymbol{\Theta}^*_\varepsilon, \tag{48}$$

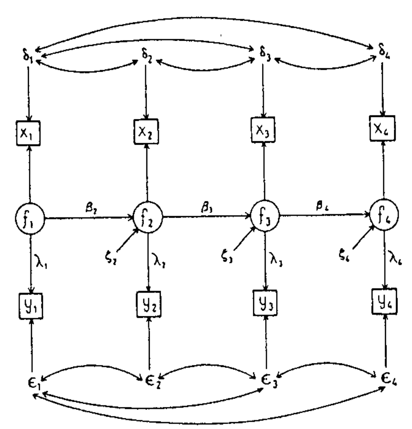

FIGURE 8. A four-wave, two-variable model with correlated errors.

where $\boldsymbol{\Phi}$, $\boldsymbol{\Theta}^*_\delta$ and $\boldsymbol{\Theta}^*_\varepsilon$ are the covariance matrices of f, δ and ε, respectively. It is seen that $\boldsymbol{D}_\lambda$ may be multiplied by a non-zero scalar, $\boldsymbol{\Phi}$ divided by the same scalar, and with $\boldsymbol{\Theta}^*_\delta$ and $\boldsymbol{\Theta}^*_\varepsilon$ properly adjusted, $\boldsymbol{\Sigma}$ will not change. Hence the model is not identified. One restriction is needed but there does not seem to be any natural way to choose it. We shall therefore consider three alternative models which are all identified. These models all represent different specification of the correlation structures for the errors in $\boldsymbol{\delta}$ and $\boldsymbol{\varepsilon}$ as follows:

Figure 9a: The errors are uncorrelated.
Figure 9b: The errors have one common factor.
Figure 9c: The errors have a simplex structure.

In all three models the covariance matrix $\boldsymbol{\Phi}$ is generated by a simplex (or first-order autoregressive) model,

$$f_{i+1} = \beta_{i+1} f_i + \zeta_{i+1}, \qquad i = 1, 2, 3.$$

This implies that

$$\boldsymbol{\Phi} = \begin{bmatrix} \phi_1 & & & \\ \beta_2\phi_1 & \phi_2 & & \\ \beta_2\beta_3\phi_1 & \beta_3\phi_2 & \phi_3 & \\ \beta_2\beta_3\beta_4\phi_1 & \beta_3\beta_4\phi_2 & \beta_4\phi_3 & \phi_4 \end{bmatrix}. \tag{49}$$

where, as before, $\phi_i = \mathrm{var}\,(f_i)$, $i = 1, 2, 3, 4$. We now examine the three models in turn.

Model 9a

To show that the model in Figure 9a is identified we first show that $\boldsymbol{\Phi}$ is identified even if it is unrestricted. In Model 9a, $\boldsymbol{\Theta}^*_\delta$ and $\boldsymbol{\Theta}^*_\varepsilon$ are diagonal, and are denoted by $\boldsymbol{\Theta}^2_\delta$ and $\boldsymbol{\Theta}^2_\varepsilon$, respectively. We can use the off-diagonal part of (46) to solve for

$$\phi_{ij} = \sigma^{(xx)}_{ij}, \qquad i \neq j.$$

Then, from (47) we have, for a given i,

$$\sigma^{(yx)}_{ij} = \lambda_i \phi_{ij}, \qquad j \neq i,$$

which represents three equations in λ_i. Hence each λ_i is overidentified with two restrictions. With $\boldsymbol{D}_\lambda$ determined, we can now use the diagonal of (47) to determine the diagonal elements of $\boldsymbol{\Phi}$, i.e.,

$$\sigma^{(yx)}_{ii} = \lambda_i \phi_{ii}.$$

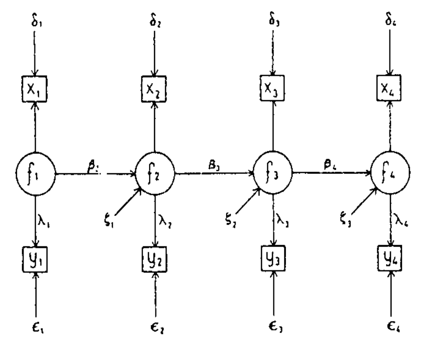

FIGURE 9a. A four-wave, two-variable model with uncorrelated errors.

The diagonal matrices $\boldsymbol{\Theta}_\delta$ and $\boldsymbol{\Theta}_\epsilon$ are then determined from the diagonal parts of (46) and (48). The off-diagonal part of (48) then yields six overidentifying equations for $\boldsymbol{D}_\lambda$ and ϕ_{ij} $(i \neq j)$. This analysis shows that each diagonal element in $\boldsymbol{\Theta}_\delta$, $\boldsymbol{\Theta}_\epsilon$, and $\boldsymbol{\Phi}$ is determined by one equation whereas the λ_i $(i = 1, 2, 3, 4)$ and the ϕ_{ij} $(i \neq j)$ are determined with fourteen overidentifying restrictions. Finally (49) is used to determine β_2, β_3 and β_4 from ϕ_{21}, ϕ_{32} and ϕ_{43}, respectively. Elements ϕ_{ij} with $i - j > 1$ yield three additional overidentifying restrictions. Hence the total number of overidentifying restrictions is 17, which checks with the number of degrees of freedom computed as number of distinct elements in $\boldsymbol{\Sigma}$ (36) minus the number of free parameters (4ϕ's, 3β's, 4λ's, 4 in $\boldsymbol{\Theta}_\delta$ and 4 in $\boldsymbol{\Theta}_\epsilon$ for a total of 19). In the general case of $m \geq 2$ occasions the number of degrees of freedom will be $2m^2 - 4m + 1$.

Model 9a can be put into the LISREL format with no $\boldsymbol{x}$. The measurement model is

$$\begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{pmatrix} = \begin{pmatrix} \boldsymbol{I} \\ \boldsymbol{D}_\lambda \end{pmatrix} \boldsymbol{f} + \boldsymbol{e},$$

and the structural model is

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ -\beta_2 & 1 & 0 & 0 \\ 0 & -\beta_3 & 1 & 0 \\ 0 & 0 & -\beta_4 & 1 \end{bmatrix} \begin{pmatrix} f_1 \\ f_2 \\ f_3 \\ f_4 \end{pmatrix} = \begin{pmatrix} \zeta_1 \\ \zeta_2 \\ \zeta_3 \\ \zeta_4 \end{pmatrix}$$

As before, LISREL treats $\psi_i = \mathrm{var}\,(\zeta_i)$ as primary parameters rather than $\phi_i = \mathrm{var}\,(f_i)$. Estimates of ϕ_i are obtained as a by-product:

$$\phi_1 = \psi_1,$$
$$\phi_i = \psi_i + \beta_i^2 \phi_{i-1}, \qquad i = 2, 3, 4.$$

Model 9b

In Figure 9b it is assumed that correlations between the errors are accounted for by a common factor. The common factors ξ and η are *test-specific* in contrast to the factors f_1, f_2, f_3 and f_4, which are *occasion-specific* [Jöreskog (1970a)]. The test-specific factors ξ and η are assumed to be uncorrelated, uncorrelated with f, δ and ε, and scaled to unit variance.

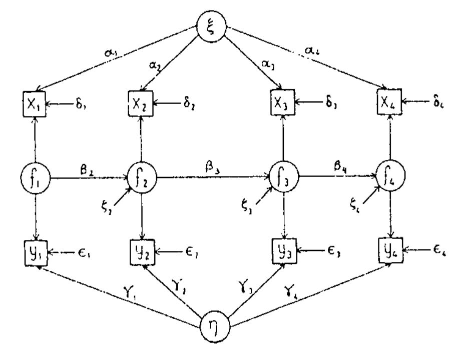

FIGURE 9b. A four-wave, two-variable model with test-specific factors.

The measurement model is

$$\boldsymbol{x} = \boldsymbol{f} + \boldsymbol{\alpha}\xi + \boldsymbol{\delta}, \tag{50}$$
$$\boldsymbol{y} = \boldsymbol{D}_\lambda \boldsymbol{f} + \boldsymbol{\gamma}\eta + \boldsymbol{\varepsilon}, \tag{51}$$

where $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$ are factor loadings relating the observed variables $\boldsymbol{x}$ and $\boldsymbol{y}$ to the test-specific factors ξ and η, respectively. The blocks of the covariance matrix of the observed variables are

$$\boldsymbol{\Sigma}_{xx} = \boldsymbol{\Phi} + \boldsymbol{\alpha\alpha}' + \boldsymbol{\Theta}_\delta^2, \tag{52}$$
$$\boldsymbol{\Sigma}_{yx} = \boldsymbol{D}_\lambda \boldsymbol{\Phi}, \tag{53}$$
$$\boldsymbol{\Sigma}_{yy} = \boldsymbol{D}_\lambda \boldsymbol{\Phi} \boldsymbol{D}_\lambda + \boldsymbol{\gamma\gamma}' + \boldsymbol{\Theta}_\varepsilon^2. \tag{54}$$

The identification problem for Model 9b is slightly more complicated than for Model 9a. First we show that $\boldsymbol{D}_\lambda$ and $\boldsymbol{\Phi}$ can be determined from (53) alone. Since for $i \neq j$,

$$\sigma_{ij}^{(yx)} = \lambda_i \phi_{ij},$$
$$\sigma_{ji}^{(yx)} = \lambda_j \phi_{ji},$$

we have

$$\lambda_i / \lambda_j = \sigma_{ij}^{(yx)} / \sigma_{ji}^{(yx)}.$$

Hence the ratios λ_1/λ_2, λ_1/λ_3, λ_1/λ_4, λ_2/λ_3, λ_2/λ_4, λ_3/λ_4 are determined. This yields six equations in four unknowns which can be solved with two overidentifying restrictions on the λ's. With $\boldsymbol{D}_\lambda$ determined, $\boldsymbol{\Phi}$ is given by (53) as

$$\boldsymbol{\Phi} = \boldsymbol{\Sigma}_{yx} \boldsymbol{D}_\lambda^{-1}.$$

From the off-diagonal elements of $\boldsymbol{\Sigma}_{xx} - \boldsymbol{\Phi} = \boldsymbol{\Sigma}^*$, say, we have

$$\sigma_{ij}^* = \alpha_i \alpha_j, \qquad i \neq j,$$

which can be used to determine α_1 from

$$\alpha_1^2 = (\sigma_{21}^* \sigma_{31}^*)/\sigma_{32}^* = (\sigma_{21}^* \sigma_{41}^*)/\sigma_{42}^* = (\sigma_{31}^* \sigma_{41}^*)/\sigma_{43}^*,$$

and similarly for α_2, α_3 and α_4. This yields six equations in four unknowns which can be solved with two overidentifying restrictions on the α's. Next $\boldsymbol{\Theta}_\delta$ is determined from the diagonal elements of (52). Finally, $\boldsymbol{\gamma}$ and $\boldsymbol{\Theta}_\epsilon$ are determined in a parallel way from (54). This analysis shows that if $\boldsymbol{\Phi}$ were free the model would be overidentified by six restrictions. The β's are determined by (49) as before, yielding three additional overidentifying restrictions. Hence, the total number of overidentifying restrictions is 9 which is the degrees of freedom of the model. In the general case with $m \geq 3$ occasions, the degrees of freedom will be $2m^2 - 6m + 1$.

Model 9a may be interpreted as the special case of Model 9b which arises when both $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$ are zero. The hypothesis $\boldsymbol{\alpha} = \boldsymbol{0}$ and $\boldsymbol{\gamma} = \boldsymbol{0}$ may thus be tested with eight degrees of freedom.

Model 9b can be put into the LISREL format using the no-$\boldsymbol{x}$ option.

The parameter matrices are specified as

$$\boldsymbol{\Lambda}_y = \begin{bmatrix} 1 & 0 & 0 & 0 & \alpha_1 & 0 \\ 0 & 1 & 0 & 0 & \alpha_2 & 0 \\ 0 & 0 & 1 & 0 & \alpha_3 & 0 \\ 0 & 0 & 0 & 1 & \alpha_4 & 0 \\ \lambda_1 & 0 & 0 & 0 & 0 & \gamma_1 \\ 0 & \lambda_2 & 0 & 0 & 0 & \gamma_2 \\ 0 & 0 & \lambda_3 & 0 & 0 & \gamma_3 \\ 0 & 0 & 0 & \lambda_4 & 0 & \gamma_4 \end{bmatrix},$$

$$\boldsymbol{B} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ -\beta_2 & 1 & 0 & 0 & 0 & 0 \\ 0 & -\beta_3 & 1 & 0 & 0 & 0 \\ 0 & 0 & -\beta_4 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix},$$

$$\boldsymbol{\Psi} = \operatorname{diag}(\psi_1, \psi_2, \psi_3, \psi_4, 1, 1),$$

where $\psi_i = \operatorname{var}(\zeta_i)$, $i = 1, 2, 3, 4$, $\boldsymbol{\Theta}_\delta$ and $\boldsymbol{\Theta}_\epsilon$ are as before.

Model 9c

In Figure 9c it is assumed that the errors have a simplex structure of the

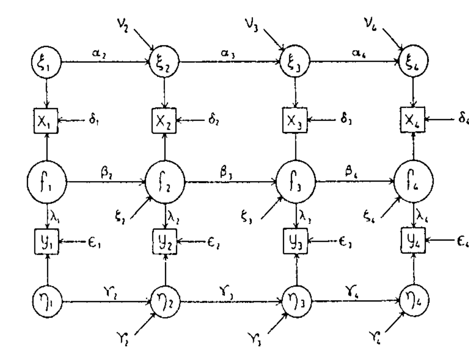

FIGURE 9c. A four-wave, two-variable model with simplex errors not identified.

type introduced in Section 4.1. The equations are

$$\boldsymbol{x} = \boldsymbol{f} + \boldsymbol{\xi} + \boldsymbol{\delta}, \tag{55}$$

$$\boldsymbol{y} = \boldsymbol{D}_\lambda \boldsymbol{f} + \boldsymbol{\eta} + \boldsymbol{\varepsilon}, \tag{56}$$

where

$$\xi_i = \alpha_i \xi_{i-1} + \nu_i,$$

$$\eta_i = \gamma_i \eta_{i-1} + \upsilon_i, \qquad i = 2, 3, 4. \tag{57}$$

The simplex variables $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$ are assumed to be uncorrelated with $\boldsymbol{\delta}$, $\boldsymbol{\varepsilon}$ and $\boldsymbol{f}$, and the residuals ν_i and υ_i are uncorrelated with each other and with ν_j, υ_j, ξ_j and η_j for $j < i$.

The covariance matrix of $\boldsymbol{\Sigma}$ is

$$\boldsymbol{\Sigma}_{xx} = \boldsymbol{\Phi} + \boldsymbol{\Sigma}_{\xi\xi} + \boldsymbol{\Theta}_\delta^2, \tag{58}$$

$$\boldsymbol{\Sigma}_{yx} = \boldsymbol{D}_\lambda \boldsymbol{\Phi}, \tag{59}$$

$$\boldsymbol{\Sigma}_{yy} = \boldsymbol{D}_\lambda \boldsymbol{\Phi} \boldsymbol{D}_\lambda + \boldsymbol{\Sigma}_{\eta\eta} + \boldsymbol{\Theta}_\varepsilon^2, \tag{60}$$

where $\boldsymbol{\Sigma}_{\xi\xi}$ and $\boldsymbol{\Sigma}_{\eta\eta}$ are the covariance matrices of $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$.

Equation (59) determines $\boldsymbol{\Phi}$ and $\boldsymbol{D}_\lambda$ as before with two overidentifying restrictions. As in Section 4.1, the covariance matrix of $\boldsymbol{\xi} + \boldsymbol{\delta}$ is

$$\boldsymbol{\Sigma}_{\xi\xi} + \boldsymbol{\Theta}_\delta^2 = \begin{bmatrix} \sigma_{\xi_1}^2 + \theta_{\delta_1}^2 & & & \\ \alpha_2\sigma_{\xi_1}^2 & \sigma_{\xi_2}^2 + \theta_{\delta_2}^2 & & \\ \alpha_2\alpha_3\sigma_{\xi_1}^2 & \alpha_3\sigma_{\xi_2}^2 & \sigma_{\xi_3}^2 + \theta_{\delta_3}^2 & \\ \alpha_2\alpha_3\alpha_4\sigma_{\xi_1}^2 & \alpha_3\alpha_4\sigma_{\xi_2}^2 & \alpha_4\sigma_{\xi_3}^2 & \sigma_{\xi_4}^2 + \theta_{\delta_4}^2 \end{bmatrix},$$

where $\sigma_{\xi_i}^2 = \text{var}(\xi_i)$ and $\theta_{\delta_i}^2 = \text{var}(\delta_i)$, $i = 1, 2, 3, 4$. Since $\boldsymbol{\Sigma}_{\xi\xi}$ and $\boldsymbol{\Theta}_\delta$ enter only into $\boldsymbol{\Sigma}_{xx}$ it is apparent that $\sigma_{\xi_1}^2$, α_2, σ_{δ_1}, $\theta_{\xi_4}^2$ and θ_{δ_4} are not identified. If one fixes $\sigma_{\xi_1}^2$ and $\sigma_{\xi_4}^2$ to unity, all the other parameters are identified with one overidentifying restriction. The situation is exactly the same as for the single simplex in Section 4.1. Similarly, fixing $\sigma_{\eta_1}^2$ and $\sigma_{\eta_4}^2$ to unity identifies all the parameters in $\boldsymbol{\Sigma}_{\eta\eta} + \boldsymbol{\Theta}_\varepsilon^2$ with one overidentifying restriction. The β's are determined from $\boldsymbol{\Phi}$ in (49) with three overidentifying restrictions as before. Altogether there will be seven overidentifying restrictions in Model 9c.

The indeterminacy in the two simplex models postulated for $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$ means that all that one can do is to estimate the path from ξ_2 to x_1 and the path from ξ_3 to x_4. Similarly one can only estimate the paths from η_2 to y_1 and from η_3 to y_4. For this reason the model in Figure 9c is modified as in Figure 10, i.e., for $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$ the end variables are eliminated. The model in

Figure 10 is an identified reparameterization of that in Figure 9c. However, with four occasions there are only two occasions between the first and the last occasion, so the simplex concept no longer applies. Still the model in Figure 10 may be interpreted as a generalization of Model 9b from one test-specific factor to two test-specific factors for each set of variables. Model 9b is a special case of the model in Figure 10 with $\text{var}(\nu) = \text{var}(\upsilon) = 0$. The hypothesis $\text{var}(\nu) = \text{var}(\upsilon) = 0$ may be tested with two degrees of freedom. As illustrated in Figure 11 the simplex concept for the errors would be meaningful when there are more than four occasions.

The model in Figure 10 can be put into the LISREL format. We do so in a way that facilitates generalization to more than four occasions. The measurement model is

$$\begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{pmatrix} = \begin{pmatrix} \boldsymbol{I} & \boldsymbol{P}_\alpha & \boldsymbol{0} \\ \boldsymbol{D}_\lambda & \boldsymbol{0} & \boldsymbol{P}_\gamma \end{pmatrix} \begin{pmatrix} \boldsymbol{f} \\ \boldsymbol{\xi} \\ \boldsymbol{\eta} \end{pmatrix} + \begin{pmatrix} \boldsymbol{\delta} \\ \boldsymbol{\varepsilon} \end{pmatrix},$$

where

$$\boldsymbol{P}_\alpha = \begin{bmatrix} \alpha_2 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & \alpha_4 \end{bmatrix}, \qquad \boldsymbol{P}_\gamma = \begin{bmatrix} \gamma_2 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & \gamma_4 \end{bmatrix},$$

$\boldsymbol{\xi} = (\xi_2, \xi_3)'$ and $\boldsymbol{\eta} = (\eta_2, \eta_3)'$. The structural model is

$$\begin{bmatrix} \boldsymbol{B}_\beta & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{B}_\alpha & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{B}_\gamma \end{bmatrix} \begin{pmatrix} \boldsymbol{f} \\ \boldsymbol{\xi} \\ \boldsymbol{\eta} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\zeta} \\ \boldsymbol{\nu} \\ \boldsymbol{\upsilon} \end{pmatrix},$$

where

$$\boldsymbol{B}_\beta = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -\beta_2 & 1 & 0 & 0 \\ 0 & -\beta_3 & 1 & 0 \\ 0 & 0 & -\beta_4 & 1 \end{bmatrix}, \quad \boldsymbol{B}_\alpha = \begin{pmatrix} 1 & 0 \\ -\alpha_3 & 1 \end{pmatrix}, \quad \boldsymbol{B}_\gamma = \begin{pmatrix} 1 & 0 \\ -\gamma_3 & 1 \end{pmatrix},$$

$\boldsymbol{\zeta} = (\zeta_1 \equiv f_1, \zeta_2, \zeta_3, \zeta_4)'$, $\boldsymbol{\nu} = (\nu_2 \equiv \xi_2, \nu_3)'$, and $\boldsymbol{\upsilon} = (\upsilon_2 \equiv \eta_2, \upsilon_3)'$. The LISREL program gives estimates of λ_1, λ_2, λ_3, λ_4, β_2, β_3, β_4, α_2, α_3, α_4, γ_2, γ_3, γ_4 and the variances of ζ_1, ζ_2, ζ_3, ζ_4, ν_2, ν_3, υ_2, υ_3. The estimates of the

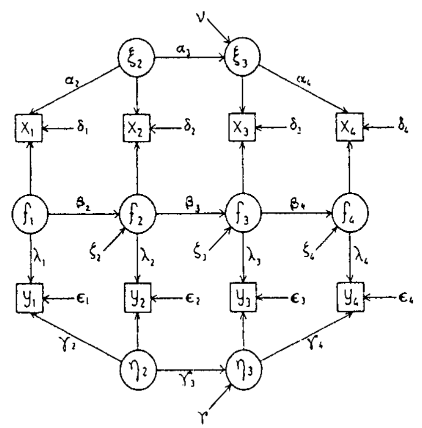

FIGURE 10. A four-wave, two-variable model with simplex errors identified.

variances of f_i, ξ_i and η_i may be recovered as

$$\mathrm{var}\,(f_1) = \mathrm{var}\,(\zeta_1),$$
$$\mathrm{var}\,(f_i) = \beta_i^2\,\mathrm{var}\,(f_{i-1}) + \mathrm{var}\,(\zeta_i), \qquad i = 2, 3, 4,$$
$$\mathrm{var}\,(\xi_2) = \mathrm{var}\,(\nu_2),$$
$$\mathrm{var}\,(\xi_3) = \alpha_3^2\,\mathrm{var}\,(\xi_2) + \mathrm{var}\,(\nu_3),$$
$$\mathrm{var}\,(\eta_2) = \mathrm{var}\,(\upsilon_2),$$
$$\mathrm{var}\,(\eta_3) = \gamma_3^2\,\mathrm{var}\,(\eta_2) + \mathrm{var}\,(\upsilon_3).$$

4.3. *Mathematical aptitudes and achievements*

To illustrate the theory we draw on a large growth study conducted at Educational Testing Service [Anderson and Maier (1963), Hilton (1969)]. A nationwide sample of fifth graders was tested in 1961 and then again in 1963, 1965 and 1967 as seventh, ninth and eleventh graders, respectively. The test scores included the verbal (SCATV) and quantitative (SCATQ) parts of SCAT (Scholastic Aptitude Test) and achievement tests in mathematics (MATH), science (SCI), social studies (SS), reading

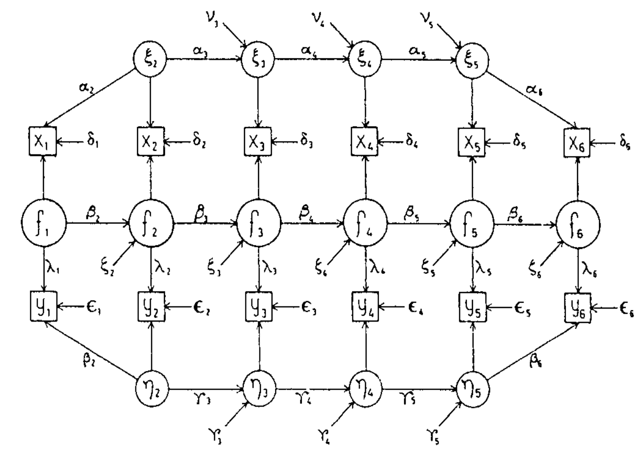

FIGURE 11. A six-wave, two-variable model with simplex errors.

(READ) and writing (WRIT). The analyses which we present are based on a subsample of 383 girls for which complete data were available for all the grades 5, 7, 9 and 11. In this section we analyze the data on MATH and SCATQ from all four occasions under Models 9a, 9b and 9c.

The maximum-likelihood estimates of the parameters are shown in Table 2. The χ^2-values clearly reject Model 9a in favor of Model 9b or 9c, but it is not so easy to discriminate between 9b and 9c. The difference in χ^2 between Model 9b and 9c is 9.85 with 2 degrees of freedom, which is significant at the 1% level. Hence it seems best to use Model 9c rather than 9b.

For interpretive purposes it may be preferable to work with transformations of the LISREL parameters. For Models 9b and 9c, Table 3 gives the estimates of the factor variances and of the squared correlations R_i^2 between f_i and f_{i-1}; Table 4 gives the covariance matrices of the errors $\boldsymbol{\varepsilon}^* = \boldsymbol{y} - \boldsymbol{D}_y\boldsymbol{f}$ and $\boldsymbol{\delta}^* = \boldsymbol{x} - \boldsymbol{f}$ (i.e., the partial covariance matrices of $\boldsymbol{y}$ and $\boldsymbol{x}$ after elimination of $\boldsymbol{f}$). Table 5 gives the corresponding correlation matrices.

The squared correlations R_i^2 in Table 3 are quite high, indicating a very high stability of the quantitative factor over time. Table 4 shows that covariation among the errors is present for the SCATQ tests to a larger extent than for the MATH tests. Table 5 shows that the residual

TABLE 2

Parameter estimates for the models in Figure 9a–c with x = MATH and y = SCATQ, N = 383.[a]

Parameter	Model 9a	Model 9b	Model 9c
λ_1	0.88	0.85	0.83
λ_2	1.13	1.06	1.05
λ_3	1.23	1.14	1.14
λ_4	1.31	1.21	1.21
β_2	1.22	1.22	1.20
β_3	1.01	1.00	1.01
β_4	1.06	1.05	1.05
$\sigma^2_{\eta_1}$	55.25	57.31	58.20
$\sigma^2_{\zeta_2}$	9.28	10.93	12.65
$\sigma^2_{\zeta_3}$	10.29	13.74	11.06
$\sigma^2_{\zeta_4}$	2.37	5.79	6.51
σ_{δ_1}	6.19	5.94	5.82
σ_{δ_2}	6.20	5.84	5.80
σ_{δ_3}	5.77	2.29	2.79
σ_{δ_4}	7.55	6.83	6.85
σ_{ϵ_1}	4.53	4.47	4.45
σ_{ϵ_2}	6.18	5.83	4.21
σ_{ϵ_3}	7.40	6.87	6.88
σ_{ϵ_4}	7.14	6.51	6.08
χ^2	72.49	23.18	13.33
d.f.	17	9	7

[a]Additional parameter estimates

Model 9b		*Model 9c*	
$\alpha_1 = 0.97$,	$\gamma_1 = 1.48$	$\alpha_2 = 5.89$,	$\gamma_2 = 0.32$
$\alpha_2 = 0.26$,	$\gamma_2 = 3.51$	$\alpha_3 = 16.06$,	$\gamma_3 = 0.45$
$\alpha_3 = 4.59$,	$\gamma_3 = 4.66$	$\alpha_4 = -0.20$,	$\gamma_4 = 1.15$
$\alpha_4 = -0.92$,	$\gamma_4 = 4.90$	$\sigma^2_{\zeta_1} = 0.04$,	$\sigma^2_u = 7.55$
		$\sigma^2_{\eta_2} = 31.05$,	$\sigma^2_v = 15.64$

TABLE 3

Factor variances and squared multiple correlations for Models 9b and 9c.

Model 9b		Model 9c	
$\sigma^2_{\eta_i}$	R^2_i	$\sigma^2_{\eta_i}$	R^2_i
57.31	-	58.20	-
96.23	0.886	96.46	0.869
109.97	0.875	109.46	0.899
127.03	0.954	127.19	0.949

TABLE 4

Covariance matrices of ϵ^* and δ^* for Models 9b and 9c.

	Model 9b	Model 9c
ϵ^*	$\begin{bmatrix} 22.17 & & & \\ 5.19 & 46.31 & & \\ 6.90 & 16.36 & 68.91 & \\ 7.25 & 17.20 & 22.83 & 66.39 \end{bmatrix}$	$\begin{bmatrix} 22.98 & & & \\ 9.94 & 48.77 & & \\ 4.47 & 13.97 & 69.26 & \\ 5.14 & 16.07 & 25.22 & 65.97 \end{bmatrix}$
δ^*	$\begin{bmatrix} 36.22 & & & \\ 0.25 & 34.17 & & \\ 4.45 & 1.19 & 26.31 & \\ -0.89 & -0.24 & -4.22 & 47.49 \end{bmatrix}$	$\begin{bmatrix} 35.26 & & & \\ 0.24 & 33.68 & & \\ 3.78 & 0.64 & 25.65 & \\ -0.76 & -0.13 & -3.57 & 47.35 \end{bmatrix}$

TABLE 5

Correlation matrices of ϵ^* and δ^* for Models 9b and 9c.

	Model 9b	Model 9c
ϵ^*	$\begin{bmatrix} 1.00 & & & \\ 0.16 & 1.00 & & \\ 0.18 & 0.29 & 1.00 & \\ 0.19 & 0.31 & 0.34 & 1.00 \end{bmatrix}$	$\begin{bmatrix} 1.00 & & & \\ 0.30 & 1.00 & & \\ 0.11 & 0.24 & 1.00 & \\ 0.13 & 0.28 & 0.37 & 1.00 \end{bmatrix}$
δ^*	$\begin{bmatrix} 1.00 & & & \\ 0.01 & 1.00 & & \\ 0.14 & 0.04 & 1.00 & \\ -0.02 & -0.01 & -0.12 & 1.00 \end{bmatrix}$	$\begin{bmatrix} 1.00 & & & \\ 0.01 & 1.00 & & \\ 0.13 & 0.02 & 1.00 & \\ -0.02 & -0.00 & -0.10 & 1.00 \end{bmatrix}$

correlations among the ϵ's are in general higher than those among the δ's; the latter are indeed very small. Hence the model accounts for the intercorrelations among the MATH tests much better than the intercorrelations among the SCATQ tests.

5. Multi-wave, multi-variable models

5.1. A general model for longitudinal data

In this section we develop a general model for longitudinal data which generalizes the models of Section 4 to the situation when several (more than two) variables are involved at several occasions and also allows for background variables as in Section 3.

We assume that p_t variables $y'_t = (y_{1t}, y_{2t}, \ldots, y_{p_t t})$ are measured at

occasion t ($t = 1, 2, \ldots, m$). At each occasion $\mathbf{y}_t$ has a factor structure with m_t correlated common factors $\boldsymbol{\eta}'_t = (\eta_{1t}, \eta_{2t}, \ldots, \eta_{m_t,t})$ so that

$$\mathbf{y}_t = \boldsymbol{\Lambda}_{yt}\boldsymbol{\eta}_t + \boldsymbol{\varepsilon}_t, \tag{61}$$

where $\boldsymbol{\varepsilon}_t$ is the vector of unique factors and $\boldsymbol{\Lambda}_{yt}$ is the factor-loading matrix of order $p_t \times m_t$.

In addition to the variables $\mathbf{y}_t$, we assume that q background variables $\mathbf{x}' = (x_1, x_2, \ldots, x_q)$ are measured, representing characteristics and conditions existing before the first occasion and assumed to influence all the $\mathbf{y}_t$ variables. The background variables are in general assumed to be fallible measures having a factor structure with common factors $\boldsymbol{\xi}' = (\xi_1, \xi_2, \ldots, \xi_n)$ so that

$$\mathbf{x} = \boldsymbol{\Lambda}_x\boldsymbol{\xi} + \boldsymbol{\delta}, \tag{62}$$

where $\boldsymbol{\delta}$ is the vector of unique factors and $\boldsymbol{\Lambda}_x$ is the factor-loading matrix of order $q \times n$.

The structural equations connecting the $\boldsymbol{\eta}$'s and $\boldsymbol{\xi}$'s are assumed to be

$$\boldsymbol{\eta}_1 = \boldsymbol{A}_1\boldsymbol{\xi} + \boldsymbol{\zeta}_1, \tag{63}$$

$$\boldsymbol{\eta}_t = \boldsymbol{A}_t\boldsymbol{\xi} + \boldsymbol{B}_t\boldsymbol{\eta}_{t-1} + \boldsymbol{\zeta}_t, \qquad t = 2, \ldots, m, \tag{64}$$

where $\boldsymbol{A}_t$ is a matrix of order $m_t \times n$, and $\boldsymbol{B}_t$ is a matrix of order $m_t \times m_{t-1}$. The residual vectors $\boldsymbol{\zeta}'_t = (\zeta_{1t}, \zeta_{2t}, \ldots, \zeta_{m_t,t})$ are assumed to be correlated within occasions but uncorrelated between occasions.

Model (61)–(64) can be put into the general LISREL format. To see this, we illustrate with $m = 4$ occasions and write (61) as

$$\begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \mathbf{y}_3 \\ \mathbf{y}_4 \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Lambda}_{y1} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Lambda}_{y2} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{\Lambda}_{y3} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{\Lambda}_{y4} \end{pmatrix} \begin{pmatrix} \boldsymbol{\eta}_1 \\ \boldsymbol{\eta}_2 \\ \boldsymbol{\eta}_3 \\ \boldsymbol{\eta}_4 \end{pmatrix} + \begin{pmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \\ \boldsymbol{\varepsilon}_3 \\ \boldsymbol{\varepsilon}_4 \end{pmatrix}, \tag{65}$$

and (63) and (64) as

$$\begin{pmatrix} \boldsymbol{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ -\boldsymbol{B}_2 & \boldsymbol{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -\boldsymbol{B}_3 & \boldsymbol{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & -\boldsymbol{B}_4 & \boldsymbol{I} \end{pmatrix} \begin{pmatrix} \boldsymbol{\eta}_1 \\ \boldsymbol{\eta}_2 \\ \boldsymbol{\eta}_3 \\ \boldsymbol{\eta}_4 \end{pmatrix} = \begin{pmatrix} \boldsymbol{A}_1 \\ \boldsymbol{A}_2 \\ \boldsymbol{A}_3 \\ \boldsymbol{A}_4 \end{pmatrix} \boldsymbol{\xi} + \begin{pmatrix} \boldsymbol{\zeta}_1 \\ \boldsymbol{\zeta}_2 \\ \boldsymbol{\zeta}_3 \\ \boldsymbol{\zeta}_4 \end{pmatrix}. \tag{66}$$

Particular cases may be useful in various situations. If each of the observed variables measures a different characteristic, trait, or construct,

the assumptions that $\mathbf{y}$ and $\mathbf{x}$ have a common factor structure may not be reasonable. Each variable may represent its own factor, i.e., we have only a single indicator of each factor. In this case it is usually not possible to estimate the unique factor or the measurement error in the variables. This case is handled by specifying $m_t = p_t$, $\boldsymbol{\Lambda}_{yt} = \boldsymbol{I}$, $\boldsymbol{\varepsilon}_t = \mathbf{0}$ ($t = 1, 2, \ldots, m$), $\boldsymbol{\Lambda}_x = \boldsymbol{I}$, and $\boldsymbol{\delta} = \mathbf{0}$. The structural equations (66) will then be defined in terms of the observed variables.

Another special case is when $\boldsymbol{A}_2 = \boldsymbol{A}_3 = \cdots = \boldsymbol{A}_m = \mathbf{0}$. Then the background variables $\boldsymbol{\xi}$ affect $\boldsymbol{\eta}_1$ directly and $\boldsymbol{\eta}_2, \boldsymbol{\eta}_3, \ldots, \boldsymbol{\eta}_m$ only indirectly via $\boldsymbol{\eta}_1$. Assumptions of this kind may be tested on the basis of available data.

A third special case is when measurements on background variables are not available. Then (62) and (63) are omitted and (64) is replaced by

$$\boldsymbol{\eta}_t = \boldsymbol{B}_t\boldsymbol{\eta}_{t-1} + \boldsymbol{\zeta}_t, \qquad t = 2, \ldots, m, \tag{67}$$

which is the direct generalization of the simplex model (39) to the multivariate case. This model is handled by using the no-$\mathbf{x}$ option of the LISREL program.

In some longitudinal studies it may be reasonable to assume that all the ε's are uncorrelated, corresponding to Model 9a of Section 4. However, as demonstrated previously, if the same measurements are used at each occasion, the corresponding ε's will tend to correlate, and in order to get consistent estimates of the $\boldsymbol{B}$'s, one must include such correlations in the model. If background variables are included in the model it is possible to estimate the intercorrelations among the ε's for the same variables directly as in Section 3.5. When background variables are not included, on the other hand, one must specify some structure for these intercorrelations, such as Model 9b or 9c of Section 4.2.

Each of the matrices $\boldsymbol{\Lambda}_{yt}$, $\boldsymbol{\Lambda}_x$, $\boldsymbol{A}_t$ and $\boldsymbol{B}_t$ may contain fixed, free, and constrained parameters. For the model to be identified it is necessary that there be an adequate number of fixed zeros in these matrices.

5.2. Building a model for aptitude and achievement measures

In longitudinal studies there are often many variables involved and a model of the kind introduced in the preceding subsection can get very complicated and involve many parameters, perhaps several hundreds. The estimation and testing of such a model with the LISREL program may present considerable difficulties even on today's high-speed compu-

ters. Furthermore, the model may only be tentative, or in any case, if it does not fit the data well (which is likely), one may want to modify the model during the process of analysis. In such a situation it is probably best to build up the model in steps, fitting various parts of the model separately, and only then putting the parts together into a whole model. We illustrate how this can be done using the ETS growth data from all four occasions. We should emphasize that there are a number of alternative ways in which the analysis can be done. Presumably these give essentially the same results but this needs to be investigated further.

We begin by finding the factor structure for each occasion. We postulate a two-factor solution for each occasion with two correlated factors V = Verbal Ability and Q = Quantitative Ability. The factor matrices Λ_{yt} $(t = 1, 2, 3, 4)$ are assumed to have the following pattern:

$$\begin{array}{r c} & \begin{array}{cc} Q & V \end{array} \\ \begin{array}{r} \text{MATH} \\ \text{SCI} \\ \text{SS} \\ \text{READ} \\ \text{SCATV} \\ \text{SCATQ} \end{array} & \begin{bmatrix} x & 0 \\ x & x \\ x & x \\ 0 & x \\ 0 & 1 \\ 1 & 0 \end{bmatrix} \end{array}, \tag{68}$$

where 0 is a loading of zero, 1 is a loading of one, and x is a non-zero loading to be estimated. The unit loadings merely fix the units of measurement in the factors V and Q. The number of factors and the factor pattern should be chosen such that the intercorrelations among the variables within occasions are accounted for sufficiently well. Fitting such a model to the data for each occasion separately gives the following goodness-of-fit-values

Grade 5: $\chi^2 = 13.20$ with 6 d.f.

Grade 7: $\chi^2 = 17.64$ with 6 d.f.

Grade 9: $\chi^2 = 2.62$ with 6 d.f.

Grade 11: $\chi^2 = 11.61$ with 6 d.f.

Although these do not represent very good fits, except for grade 9, we shall nevertheless retain these models as representing reasonable approximations.

In the next step we analyze all tests at all occasions simultaneously to see to what extent the occasion-specific factors V_t and Q_t at each occasion also account for intercorrelations between tests and between

occasions. Hence we specify the following pattern as in (65):

$$\begin{array}{l r c} & & \begin{array}{cccccccc} Q_5 & V_5 & Q_7 & V_7 & Q_9 & V_9 & Q_{11} & V_{11} \end{array} \\ \begin{array}{l} \\ \\ \text{Grade 5} \\ \\ \\ \\ \\ \\ \text{Grade 7} \\ \\ \\ \\ \\ \\ \text{Grade 9} \\ \\ \\ \\ \\ \\ \text{Grade 11} \\ \\ \\ \\ \end{array} & \begin{array}{r} \text{Math} \\ \text{SCI} \\ \text{SS} \\ \text{READ} \\ \text{SCATV} \\ \text{SCATQ} \\ \text{MATH} \\ \text{SCI} \\ \text{SS} \\ \text{READ} \\ \text{SCATV} \\ \text{SCATQ} \\ \text{MATH} \\ \text{SCI} \\ \text{SS} \\ \text{READ} \\ \text{SCATV} \\ \text{SCATQ} \\ \text{MATH} \\ \text{SCI} \\ \text{SS} \\ \text{READ} \\ \text{SCATV} \\ \text{SCATQ} \end{array} & \begin{bmatrix} x & 0 & & & & & & \\ x & x & & & & & & \\ x & x & & & & & & \\ 0 & x & & & & & & \\ 0 & 1 & & & & & & \\ 1 & 0 & & & & & & \\ & & x & 0 & & & & \\ & & x & x & & & & \\ & & x & x & & & & \\ & & 0 & x & & & & \\ & & 0 & 1 & & & & \\ & & 1 & 0 & & & & \\ & & & & x & 0 & & \\ & & & & x & x & & \\ & & & & x & x & & \\ & & & & 0 & x & & \\ & & & & 0 & 1 & & \\ & & & & 1 & 0 & & \\ & & & & & & x & 0 \\ & & & & & & x & x \\ & & & & & & x & x \\ & & & & & & 0 & x \\ & & & & & & 0 & 1 \\ & & & & & & 1 & 0 \end{bmatrix} \end{array} \tag{69}$$

All blank entries are zeros. At this step we let all eight factors be freely intercorrelated. This model gives $\chi^2 = 646.78$ with 216 degrees of freedom which represents a very poor fit.

Although the intercorrelations within occasions are accounted for reasonably well, the model does not account for intercorrelations of each test between occasions. This suggests that one should let the ε's be correlated for the same variable between occasions. Since background variables are not available here we shall attempt to account for these intercorrelations by introducing test-specific factors for each of the six tests (Model 9b). Denoting these test-specific factors by S_i $(i = 1, \ldots, 6)$, this amounts to adding six columns to the previous factor matrix so that the new factor-loading matrix becomes:

		Q_5	V_5	Q_7	V_7	Q_9	V_9	Q_{11}	V_{11}	S_1	S_2	S_3	S_4	S_5	S_6
Grade 5	MATH	x	0							x					
	SCI	x	x								x				
	SS	x	x									x			
	READ	0	x										x		
	SCATV	0	1											x	
	SCATQ	1	0												x
Grade 7	MATH			x	0					x					
	SCI			x	x						x				
	SS			x	x							x			
	READ			0	x								x		
	SCATV			0	1									x	
	SCATQ			1	0										x
Grade 9	MATH					x	0			x					
	SCI					x	x				x				
	SS					x	x					x			
	READ					0	x						x		
	SCATV					0	1							x	
	SCATQ					1	0								x
Grade 11	MATH							x	0	x					
	SCI							x	x		x				
	SS							x	x			x			
	READ							0	x				x		
	SCATV							0	1					x	
	SCATQ							1	0						x

The test-specific factors S_i are assumed to have unit-variance and to be mutually uncorrelated and uncorrelated with the occasion-specific factors Q_t and V_t. We still assume that the factors Q_t and V_t are freely intercorrelated. In addition to these factors there will be unique factors for each test and occasion. This model gives $\chi^2 = 313.29$ with 192 degrees of freedom, which represents a considerable improvement in fit. The estimate of the covariance matrix of the occasion-specific factors is

$$\begin{array}{c} Q_5 \\ V_5 \\ Q_7 \\ V_7 \\ Q_9 \\ V_9 \\ Q_{11} \\ V_{11} \end{array}
\begin{bmatrix}
40.40 & & & & & & & \\
53.34 & 90.34 & & & & & & \\
60.38 & 79.78 & 104.79 & & & & & \\
62.53 & 104.74 & 109.36 & 141.63 & & & & \\
64.66 & 85.20 & 112.11 & 115.66 & 142.60 & & & \\
58.74 & 99.08 & 103.16 & 134.71 & 122.35 & 139.37 & & \\
72.27 & 96.69 & 130.43 & 132.50 & 154.46 & 133.28 & 183.35 & \\
43.38 & 83.41 & 83.01 & 110.18 & 93.59 & 110.21 & 110.55 & 104.09
\end{bmatrix}$$

Finally we fit the above covariance matrix by the bivariate simplex model

$$\begin{aligned} Q_t &= \beta_{Q_t} Q_{t-2} + q_t, \\ V_t &= \beta_{V_t} V_{t-2} + v_t, \end{aligned} \qquad t = 7, 9, 11,$$

where the residuals q_t and v_t are correlated contemporaneously but uncorrelated between occasions. This can be done by performing three separate bivariate regressions with the regression matrix constrained in each case to contain two zero coefficients. Tests of these zero constraints give

Grades 5–7: $\chi^2 = 3.31$ with 2 d.f.,
Grades 7–9: $\chi^2 = 9.11$ with 2 d.f.,
Grades 9–11: $\chi^2 = 6.80$ with 2 d.f.,

which represent reasonably good fits for these parts of the model. The full set of parameter estimates is given in Table 6.

Perhaps the most notable estimates are the large loadings 5.68, −5.72 and 5.16 of SCATQ9, MATH11 and SCATQ11 on the test-specific factors. These correspond to partial correlations, after Q_t and V_t are eliminated, of 0.39, −0.43 and 0.32, respectively. The most reasonable interpretation of these correlations is that they reflect changes in the tests SCATQ and MATH from the earlier to the later grades. In the earlier grades these tests consist mainly of arithmetic items whereas in the later

TABLE 6

Estimates for a four-wave, six-variable model with a bivariate simplex, $N = 383$.[a]

Grade	Test	Q_t	V_t	S_1	S_2	S_3	S_4	S_5	S_6	Unique variances
5	MATH	1.24	0.0*	0.68						32.49
	SCI	0.07	0.97		1.99					23.57
	SS	0.12	0.90			0.78				20.96
	READ	0.0*	1.31				4.69			46.10
	SCATV	0.0*	1.00*					2.77		28.37
	SCATQ	1.00*	0.0*						2.00	19.79
7	MATH	0.97	0.0*	0.58						33.35
	SCI	0.01	0.62		2.37					23.11
	SS	−0.09	1.05			1.54				32.45
	READ	0.0*	1.07				3.71			41.13
	SCATV	0.0*	1.00*					4.67		29.47
	SCATQ	1.00*	0.0*						4.02	37.44
9	MATH	0.90	0.0*	1.11						19.86
	SCI	−0.02	0.84		3.31					25.00
	SS	0.11	0.88			2.02				41.04
	READ	0.0*	0.86				4.03			30.81
	SCATV	0.0*	1.00*					2.74		57.18
	SCATQ	1.00*	0.00*						5.68	39.15
11	MATH	0.85	0.0*	−5.72						12.08
	SCI	0.08	0.70		3.08					46.29
	SS	−0.02	1.16			2.36				27.74
	READ	0.0*	1.18				3.72			38.11
	SCATV	0.0*	1.00*					2.68		73.47
	SCATQ	1.00*	0.0*						5.16	44.52

t	β_{Vt}	β_{Qt}	var (v_t)	var (q_t)	cov (v_t, q_t)	corr (v_t, q_t)
7	1.13	0.92	13.64	23.13	16.44	0.93
9	1.03	0.78	18.42	8.38	10.36	0.83
11	1.03	1.09	11.53	23.55	8.61	0.52

$x^2 = 313.29$ with 192 degrees of freedom

[a] Estimates obtained by stepwise analysis. Asterisks denote parameter values specified by hypothesis.

grades they are made up of items measuring logical reasoning and presented in verbal form. The fact that the loading of MATH11 is negative suggests that there is a relatively large verbal component in this test which is not accounted for by the model because of the constrained zero loading of MATH11 on V_{11}.

The relative contribution of each set of factors to the total test variance is relatively easily computed from Table 6 since the three sets of factors are uncorrelated. These variance components are given in Table 7. It is seen that the occasion-specific factors account for a large proportion of the total variance, and that the test-specific variances are relatively small even when compared with the unique variances.

6. Summary and conclusion

We have described a general model LISREL involving linear structural equations among a set of true variables which cannot be directly observed but which act as common factors of a set of observed variables. By allowing the parameters of the model to be fixed, free, or constrained,

TABLE 7

Relative variance contributions of the three sets of factors in Table 6.

Grade	Test	Occasion-specific	Test-specific	Unique
5	MATH	0.654	0.005	0.342
	SCI	0.770	0.033	0.197
	SS	0.799	0.006	0.196
	READ	0.696	0.098	0.206
	SCATV	0.715	0.061	0.224
	SCATQ	0.629	0.062	0.308
7	MATH	0.744	0.003	0.254
	SCI	0.665	0.066	0.270
	SS	0.798	0.014	0.188
	READ	0.747	0.064	0.190
	SCATV	0.734	0.113	0.153
	SCATQ	0.662	0.102	0.236
9	MATH	0.845	0.009	0.146
	SCI	0.727	0.083	0.190
	SS	0.749	0.023	0.229
	READ	0.687	0.108	0.205
	SCATV	0.683	0.037	0.280
	SCATQ	0.666	0.151	0.183
11	MATH	0.746	0.186	0.069
	SCI	0.535	0.079	0.386
	SS	0.801	0.033	0.166
	READ	0.736	0.070	0.194
	SCATV	0.563	0.039	0.398
	SCATQ	0.720	0.105	0.175

great flexibility is obtained in that the general model contains a wide range of specific models useful in the social and behavioral sciences. In this paper we focused attention on models for longitudinal or panel studies of the type often conducted in education, psychology and sociology. Using the LISREL program, each model may be estimated by the maximum-likelihood method based on the assumption that the observed variables have a multinormal distribution. When the estimates have been obtained the information matrix is computed and used to determine standard errors for the estimated parameters. The over-all fit of the model may be assessed by means of a χ^2-test.

We have discussed two-wave and multi-wave models with one, two, or more observed variables on each occasion and with or without background variables. For each model we have considered the identification problem and showed how the model may be put into the LISREL format for estimation purposes. Some of the models have been illustrated on educational and sociological data.

Some points for further research on LISREL are:

(i) How serious are departures from multinormality (a) for estimation and (b) for testing goodness-of-fit?
(ii) Can the assumption of multinormality of the observed variables be replaced by some other distributional assumption which is more valid for a wide range of problems?
(iii) How can LISREL be extended to handle data from several groups?
(iv) How can LISREL be extended to handle qualitative variables?

These topics will be studied in our future research.

References

Anderson, S.B. and M.H. Maier, 1963, 34,000 pupils and how they grew, Journal of Teacher Education 14, 212–216.
Anderson, T.W., 1959, Some stochastic process models for intelligence test scores, in: K.J. Arrow, S. Karlin and P. Suppes, eds., Mathematical methods in the social sciences (Stanford University Press, Stanford, CA).
Bohrnstedt, G.W., 1969, Observations on the measurement of change, in: E.F. Borgatta, ed., Sociological methodology (Jossey-Bass, San Francisco, CA).
Corballis, M.C., 1973, A factor model for analysing change, The British Journal of Mathematical and Statistical Psychology 26, 90–97.
Duncan, O.D., 1969, Some linear models for two-wave, two-variable panel analysis, Psychological Bulletin 72, 177–182.

Duncan, O.D., 1972, Unmeasured variables in linear models for panel analysis, in: H.L. Costner, ed., Sociological methodology (Jossey-Bass, San Francisco, CA).
Fisher, F.M., 1966, The identification problem in econometrics (McGraw-Hill, New York).
Fredriksen, C.R., 1974, Models for the analysis of alternative sources of growth in correlated stochastic variables, Psychometrika 39, 223–245.
Geraci, V.J., 1974, Simultaneous equation models with measurement error, Unpublished Ph.D. Dissertation (University of Wisconsin, Madison, WI).
Guttman, L., 1954, A new approach to factor analysis: The radex, in: P.F. Lazarsfeld, ed., Mathematical thinking in the social sciences (Columbia University Press, New York) 258–348.
Heise, D.R., 1969, Separating reliability and stability in test-retest correlation, American Sociological Review 34, 93–101.
Heise, D.R., 1970, Causal inference from panel data, in: E.F. Borgatta and G.W. Bohrnstedt, eds., Sociological methodology (Jossey-Bass, San Francisco, CA) 3–27.
Hilton, T.L., 1969, Growth study annotated bibliography, Progress Report 69–11 (Educational Testing Service, Princeton, NJ).
Jöreskog, K.G., 1970a, Factoring the multitest-multioccasion correlation matrix, in: C.E. Lunneborg, ed., Current problems and techniques in multivariate psychology, Proceedings of a conference honoring Professor Paul Horst (University of Washington, Seattle, WA) 68–100.
Jöreskog, K.G., 1970b, Estimation and testing of simplex models, The British Journal of Mathematical and Statistical Psychology 23, 121–145.
Jöreskog, K.G. and M. van Thillo, 1973, LISREL – A general computer program for estimating a linear structural equation system involving multiple indicators of unmeasured variables, Research Report 73-5 (Statistics Department, Uppsala University, Uppsala).
Kenny, D.A., 1973, Cross-lagged and synchronous common factors in panel data, in: A.S. Goldberger and O.D. Duncan, eds., Structural equation models in the social sciences (Seminar Press, New York) 153–165.
Lord, F. and M. Novick, 1968, Statistical theories of mental test scores (Addison-Wesley, Reading, MA).
Schmidt, W.H. and D.E. Wiley, 1974, Analytic problems in longitudinal data, Prepared for the Conference on Methodological Concerns in Evaluational Research (Chicago, IL).
Summers, G.F., R.L. Hough, J.T. Scott and C.L. Folse, 1969, Before industrialisation: A rural social system base study, Bulletin no. 736 (Illinois Agricultural Experiment Station, University of Illinois, Urbana, IL).
Werts, C.E., K.G. Jöreskog and R.L. Linn, 1971, Comment on "The estimation of measurement error in panel data", American Sociological Review 36, 110–113.
Werts, C.E., K.G. Jöreskog and R.L. Linn, 1972, A multitrait-multimethod model for studying growth, Educational and Psychological Measurement 32, 655–678.
Werts, C.E., K.G. Jöreskog and R.L. Linn, 1973, Identification and estimation in path analysis with unmeasured variables, American Journal of Sociology 78, 1469–1484.
Werts, C.E., K.G. Jöreskog and R.L. Linn, 1976, A simplex model for analyzing academic growth, Educational and Psychological Measurement 36.
Wheaton, B., B. Muthén, D. Alwin and G. Summers, 1977, Assessing reliability and stability in panel models with multiple indicators, in: D.R. Heise, ed., Sociological methodology, in press.
Wiley, D.E., 1973, The identification problem for structural equation models with unmeasured variables, in: A.S. Goldberger and O.D. Duncan, eds., Structural equation models in the social sciences (Seminar Press, New York) 69–83.
Wiley, D.E. and J.A. Wiley, 1970, The estimation of measurement error in panel data, American Sociological Review 35, 112–117.

Reprinted from: D. V. Aigner and A. S. Goldberg, eds., *Latent Variables in Socioeconomic Models* (Amsterdam: No. Holland Publishing Co., 1977), Chapter 16.

Chapter 6

Detection of Correlated Errors in Longitudinal Data

Dag Sörbom

A study of change in ability between two occasions may employ a number of tests believed to measure the ability in question. Either the same battery of tests is used on both occasions, or equivalent forms are used. For a variety of reasons, correlations may exist between certain errors remaining after eliminating variance due to true scores, and hence the classical factor analysis model is not applicable. A procedure for detecting correlations between errors is discussed. A search strategy is proposed since, even if the number of observed variables is small, the number of possible models is very large. A computer program is described, which produces maximum-likelihood estimates for the parameters in a factor analytic model in which the error variables may be correlated.

1. Introduction

Consider the following formal description of the classical factor analysis model (cf. Anderson, 1959).

The correlations of a number of observed variables are given. We are trying to explain these correlations by a number of variates called 'factors', fewer than the number of variables. Thus, when the influence of the factors on the variables is removed, the variables are supposed to be uncorrelated; that is,

$$\operatorname{cov}(\mathbf{x}|\mathbf{f}) = \boldsymbol{\psi}, \tag{1}$$

where $\mathbf{x}$ is a random vector of observed variables, $\mathbf{f}$ is a random vector of factors and $\boldsymbol{\psi}$ is a diagonal matrix whose elements are variances of unique or residual variables assumed to be independent of $\mathbf{f}$.

In several recent papers there has been an interest in models for which this assumption, the assumption of local independence, does not hold (cf. Corballis, 1973; Costner & Schönberg, 1973; Kenny, 1973; Wiley & Hornik, 1973). In these papers, in contrast to the classical factor analysis model, there is no assumption that all error variates are uncorrelated. For example, when some traits are measured with the same tests at several occasions it may not be reasonable to assume that the measurement errors for the same tests at different occasions are uncorrelated (cf. Corballis, 1973); but, when similar tests are used to measure the same trait, a correlation between errors of measurement within occasions seems plausible. Conditions like these give rise to a great variety of different plausible models, and the search for 'the most plausible model' may be

---

Requests for reprints should be sent to Dr Dag Sörbom, Department of Statistics, University of Uppsala, P.O. Box 2300, S-750 02 Uppsala, Sweden.

hazardous. A strategy for such a search is proposed by Costner & Schönberg (1973). An alternative procedure is given in this paper.

The aim of the procedure is to guide a sequenced search for a model with an acceptable fit to data. At each stage the current model can be examined and perhaps reformulated to result in a model with more interpretable parameters. For example, an analysis might suggest a structure of the ψ matrix which is equivalent to a model with an extra factor added.

2. An Illustrative Example

The objective of the search procedure is to detect covariation between errors (measurement errors or residuals). In this section an example is given to illustrate how the procedure works.

Suppose we want to study the change in ability between two occasions for a group of individuals. Let ξ denote the random variable representing the ability at the first occasion for a random sample from the group, and η the ability for the same individuals at the second occasion. Suppose the structural equation

$$\eta = \alpha + \beta\xi + \zeta \tag{2}$$

describes the connexion between final and initial status. The residual ζ is supposed to be uncorrelated with ξ. Our main concern is the parameters α and β in (2). To estimate these we administer three tests at each occasion. Let $\mathbf{x}$ denote the measurements at the first occasion and $\mathbf{y}$ the measurements at the second. The following model for the measurements is assumed

$$\begin{aligned} x_1 &= \lambda_{11}\xi + \varepsilon_1, \\ x_2 &= \lambda_{21}\xi + \varepsilon_2, \\ x_3 &= \lambda_{31}\xi + \varepsilon_3 \end{aligned} \tag{3a}$$

and

$$\begin{aligned} y_1 &= \lambda_{12}\eta + \delta_1, \\ y_2 &= \lambda_{22}\eta + \delta_2, \\ y_3 &= \lambda_{32}\eta + \delta_3, \end{aligned} \tag{3b}$$

$\boldsymbol{\varepsilon}$ and $\boldsymbol{\delta}$ are vectors of errors, supposed to be uncorrelated with ξ and η. If we further assume that the variables in $\boldsymbol{\varepsilon}$ and $\boldsymbol{\delta}$ are uncorrelated and that

$$E(\boldsymbol{\varepsilon}) = E(\boldsymbol{\delta}) = \mathbf{0},$$

the model is a special case of that considered by Sörbom (1973).

Let us assume that x_1 and y_1 represent very similar tests administered at the two occasions. Then there may be covariation between x_1 and y_1 not only because of the covariation between the abilities at the two occasions, ξ and η, but also because of incidental features arising from the construction of the tests or because of memory effects. Thus it may not be true that $\operatorname{cov}(\varepsilon_1, \delta_1) = 0$. Assume further that the tests x_2 and x_3 measure some traits, which are additional to but are uncorrelated with ξ. Thus covariation between x_2 and x_3 for a given value of ξ would be expected, and $\operatorname{cov}(\varepsilon_2, \varepsilon_3) \neq 0$. If we assume that the errors

for y_2 and y_3 are correlated too, the measurement model can be described as in Fig. 1.

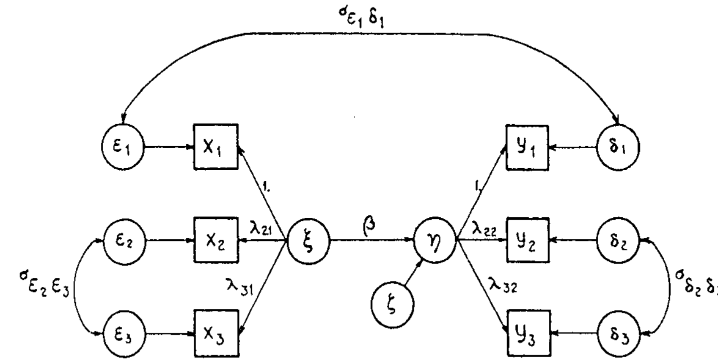

Fig. 1.—The true model. Circles denote unobservable variables and squares denote observable variables.

It should be noted that the parameters λ_{11}, λ_{21}, λ_{31} and σ_ξ^2 in (3a) cannot be identified simultaneously. We can, for example, multiply ξ by a constant and divide λ_{11}, λ_{21} and λ_{31} by the same constant without changing the measurements $\mathbf{x}$. Thus we have to fix at least one of these parameters. In the following λ_{11} is chosen to be equal to 1. Similarly λ_{12} is set equal to 1.

Let $\boldsymbol{\theta} = E\begin{pmatrix}\xi \\ \eta\end{pmatrix}$,

$$\boldsymbol{\Lambda} = \begin{bmatrix} 1 & 0 \\ \lambda_{21} & 0 \\ \lambda_{31} & 0 \\ 0 & 1 \\ 0 & \lambda_{22} \\ 0 & \lambda_{32} \end{bmatrix}, \quad \boldsymbol{\Phi} = E\left[\begin{pmatrix}\xi - \theta_1 \\ \eta - \theta_2\end{pmatrix}(\xi - \theta_1, \eta - \theta_2)\right] = \begin{bmatrix} \sigma_\xi^2 & \\ \sigma_{\xi\eta} & \sigma_\eta^2 \end{bmatrix},$$

$$\boldsymbol{\Psi} = E\left[\begin{pmatrix}\boldsymbol{\varepsilon} \\ \boldsymbol{\delta}\end{pmatrix}(\boldsymbol{\varepsilon}, \boldsymbol{\delta})\right] = \begin{bmatrix} \sigma_{\varepsilon_1}^2 & & & & & \\ 0 & \sigma_{\varepsilon_2}^2 & & & & \\ 0 & \sigma_{\varepsilon_2\varepsilon_3} & \sigma_{\varepsilon_3}^2 & & & \\ \sigma_{\varepsilon_1\delta_1} & 0 & 0 & \sigma_{\delta_1}^2 & & \\ 0 & 0 & 0 & 0 & \sigma_{\delta_2}^2 & \\ 0 & 0 & 0 & 0 & \sigma_{\delta_2\delta_3} & \sigma_{\delta_3}^2 \end{bmatrix}. \tag{4}$$

Consider the following parameter values for an artificial population:

$$\mathbf{\Lambda}_0 = \begin{bmatrix} 1\cdot0 & 0\cdot0 \\ 0\cdot8 & 0\cdot0 \\ 0\cdot5 & 0\cdot0 \\ 0\cdot0 & 1\cdot0 \\ 0\cdot0 & 0\cdot9 \\ 0\cdot0 & 0\cdot7 \end{bmatrix}, \quad \mathbf{\Phi}_0 = \begin{bmatrix} 16\cdot0 & \\ 12\cdot0 & 25\cdot0 \end{bmatrix}, \quad \sigma_{\epsilon_0}^2 = \begin{pmatrix} 4\cdot0 \\ 2\cdot56 \\ 1\cdot0 \end{pmatrix}, \quad \sigma_{\delta_0}^2 = \begin{pmatrix} 6\cdot25 \\ 5\cdot0625 \\ 3\cdot0625 \end{pmatrix}$$

and

$$\mathbf{\theta}_0 = \begin{pmatrix} 10 \\ 20 \end{pmatrix}.$$

These parameters have been chosen so that the reliability of each test equals 0·8. This means, for example, that the variance of ϵ_2 is $\lambda_{21}^2\sigma_\xi^2(1\cdot0-0\cdot8)/0\cdot8 = 2\cdot56$. It is further assumed that the population correlation between ϵ_1 and δ_1, $\rho_{\epsilon_1\delta_1}$, equals 0·5 and that $\rho_{\epsilon_2\epsilon_3}$ equals 0·3 and $\rho_{\delta_2\delta_3}$ equals 0·4.

The parameters of (2) are obtained by

$$\beta = \frac{\sigma_{\xi\eta}}{\sigma_\xi^2} = \frac{12}{16} = 0\cdot75,$$

$$\alpha = \theta_2 - \beta\theta_1 = 20-(0\cdot75)(10) = 12\cdot5,$$

and the mean values of the observed measurements are given by

$$E(\mathbf{x}) = \begin{bmatrix} \theta_1 \\ \lambda_{21}\theta_1 \\ \lambda_{31}\theta_1 \end{bmatrix} = \begin{bmatrix} 10 \\ 8 \\ 5 \end{bmatrix} \quad \text{and} \quad E(\mathbf{y}) = \begin{bmatrix} \theta_2 \\ \lambda_{22}\theta_2 \\ \lambda_{32}\theta_2 \end{bmatrix} = \begin{bmatrix} 20 \\ 18 \\ 14 \end{bmatrix}.$$

The population variance–covariance matrix for the observed measurements is given by

$$\mathbf{\Sigma}_0 = E\left[\begin{pmatrix} \mathbf{x}-E(\mathbf{x}) \\ \mathbf{y}-E(\mathbf{y}) \end{pmatrix}(\mathbf{x}-E(\mathbf{x}), \mathbf{y}-E(\mathbf{y}))\right]$$

$$= E\left[\mathbf{\Lambda}\begin{pmatrix} \xi \\ \eta \end{pmatrix} - \mathbf{\Lambda\theta} + \begin{pmatrix} \epsilon \\ \delta \end{pmatrix}\right]\left[\mathbf{\Lambda}\begin{pmatrix} \xi \\ \eta \end{pmatrix} - \mathbf{\Lambda\theta} + \begin{pmatrix} \epsilon \\ \delta \end{pmatrix}\right]' = \mathbf{\Lambda\Phi\Lambda}' + \mathbf{\Psi}$$

$$= \begin{bmatrix} \sigma_\xi^2+\sigma_{\epsilon_1}^2 & & & & & \\ \lambda_{21}\sigma_\xi^2 & \lambda_{21}^2\sigma_\xi^2+\sigma_{\epsilon_2}^2 & & & & \\ \lambda_{31}\sigma_\xi^2 & \lambda_{21}\lambda_{31}\sigma_\xi^2+\sigma_{\epsilon_2\epsilon_3} & \lambda_{31}^2\sigma_\xi^2+\sigma_{\epsilon_3}^2 & & & \\ \sigma_{\xi\eta}+\sigma_{\epsilon_1\delta_1} & \lambda_{21}\sigma_{\xi\eta} & \lambda_{31}\sigma_{\xi\eta} & \sigma_\eta^2+\sigma_{\delta_1}^2 & & \\ \lambda_{22}\sigma_{\xi\eta} & \lambda_{21}\lambda_{22}\sigma_{\xi\eta} & \lambda_{31}\lambda_{22}\sigma_{\xi\eta} & \lambda_{22}\sigma_\eta^2 & \lambda_{22}^2\sigma_\eta^2+\sigma_{\delta_2}^2 & \\ \lambda_{32}\sigma_{\xi\eta} & \lambda_{21}\lambda_{32}\sigma_{\xi\eta} & \lambda_{31}\lambda_{32}\sigma_{\xi\eta} & \lambda_{32}\sigma_\eta^2 & \lambda_{22}\lambda_{32}\sigma_\eta^2+\sigma_{\delta_2\delta_3} & \lambda_{32}^2\sigma_\eta^2+\sigma_{\delta_3}^2 \end{bmatrix}$$

$$= \begin{bmatrix} 20\cdot0000 & & & & & \\ 12\cdot8000 & 12\cdot8000 & & & & \\ 8\cdot0000 & 7\cdot3600 & 5\cdot0000 & & & \\ 14\cdot5000 & 9\cdot6000 & 6\cdot0000 & 31\cdot2500 & & \\ 10\cdot8000 & 8\cdot6400 & 5\cdot4000 & 22\cdot5000 & 25\cdot3125 & \\ 8\cdot4000 & 6\cdot7200 & 4\cdot2000 & 17\cdot5000 & 17\cdot3250 & 15\cdot3125 \end{bmatrix}. \tag{5}$$

To illustrate the procedure a sample variance–covariance matrix, **S**, and a vector of sample means, $\bar{\mathbf{z}}' = (\bar{\mathbf{x}}, \bar{\mathbf{y}})'$, was obtained by generating 1000 normal variates according to a normal distribution with mean equal to $\mathbf{\Lambda}_0\mathbf{\theta}_0$ and variance–covariance matrix equal to $\mathbf{\Sigma}_0$.

Thus it is as though for six tests we have obtained independent measurements from a sample of 1000 individuals belonging to the population of individuals under study. Three of the tests were administered at one occasion and three were administered at a later occasion. At each occasion the tests are measuring approximately the same ability. We are mainly concerned with the development of this ability between the two occasions. Because of the influence of memory effects, imperfections in the construction of the tests and similar matters, the estimates of the parameters in (2) might be distorted. The procedure to be presented is concerned with how to detect these distorting effects, and how to take them into account.

Given the data from the above simulated study, it seems natural to start with a model in which the errors are uncorrelated; that is,

$$\mathbf{\Sigma} = \mathbf{\Lambda\Phi\Lambda}' + \mathbf{\Psi}, \tag{6}$$

where $\mathbf{\Psi}$ is a diagonal matrix and $\mathbf{\Lambda}$ is structured as in (4). Maximum-likelihood

estimates of the parameters of the model may be obtained by minimizing (see Section 4)

$$f = \log|\boldsymbol{\Sigma}| + \mathrm{tr}(\mathbf{T}\boldsymbol{\Sigma}^{-1}) - \log|\mathbf{S}| - p$$

and a χ^2 measure of goodness of fit may be obtained as N times the minimum value of f (see Sörbom, 1973). The minimization of f may be done by a computer program developed by Sörbom (1974*a*). The χ^2 goodness-of-fit statistic of this model is given in Table 1, row 1. Of course, as is seen, the fit is not acceptable.

Table 1.—Goodness of Fit for the Different Models in the Simulated Example

Model	χ^2	d.f.	Probability level	χ^2 for test of the hypothesis $\psi_{ij} = 0$ (d.f. = 1)
Uncorrelated errors	216·491	12	0·000	
ψ_{32} free	163·217	11	0·000	53·274
ψ_{32}, ψ_{41} free	35·227	10	0·000	127·990
ψ_{32}, ψ_{41}, ψ_{65} free	14·430	9	0·108	20·797
ψ_{32}, ψ_{41}, ψ_{65}, ψ_{62} free	13·924	8	0·084	0·506

To improve the fit one can relax some of the restrictions imposed on the elements of either $\boldsymbol{\Lambda}$ or $\boldsymbol{\Psi}$, the elements of $\boldsymbol{\Phi}$ being already unrestricted. However, those elements in $\boldsymbol{\Lambda}$ set equal to zero reflect the basic structure of the measurements; that is, the tests are measuring one ability at each occasion. Thus the only restrictions that one would be willing to relax are the zeros in the off-diagonal elements of $\boldsymbol{\Psi}$. The question is: which? As indicated by Costner & Schönberg (1973), an examination of the differences between the observed covariances and the covariances implied by the estimated model may be misleading, when we are dealing with maximum-likelihood estimates. If, on the other hand, we had obtained the estimates by minimizing $\mathrm{tr}(\boldsymbol{\Sigma}-\mathbf{S})^2$, these differences might have been useful.

The problem can be summarized as follows. We want to find those off-diagonal elements of $\boldsymbol{\Psi}$ which are different from zero. Our metric for the goodness of fit of a model is the function f which is in principle the negative of the log-likelihood function, if the observed variables have a multinormal distribution. Thus, since we want to find the restriction which is least probable, we should relax the zero-restriction for that element which gives the largest decrease in f. This leads us to find the indices i and j such that the absolute value of $\partial f/\partial\psi_{ij}$ is greatest. In Table 2 the derivatives $\partial f/\partial\boldsymbol{\psi}$ are given for the model (6). An inspection of this table suggests that ψ_{32} is not zero; that is, there is a covariation between ε_2 and ε_3. Thus, we release the restriction $\psi_{32} = 0$ and estimate the parameters of this new model. This can be done by another computer program developed by Sörbom (1974*b*). The fit of this model, as given in Table 1, is still not acceptable, but a test of $\psi_{32} = 0$ with one degree of freedom shows that the hypotheses $\psi_{32} = 0$ is not tenable. An inspection of the derivatives $\partial f/\partial\psi_{ij}$ for this model, as given in Table 3, suggests that ψ_{41} may not be equal to zero. Thus we relax the restriction $\psi_{41} = 0$ and estimate the

Table 2.—Derivatives and Residuals for the Model with Uncorrelated Errors

Derivatives ($\partial f/\partial\boldsymbol{\psi}$)

x_1	0·000					
x_2	0·013	0·000				
x_3	0·026	−0·057	0·000			
y_1	−0·045	0·027	0·031	0·000		
y_2	0·022	−0·012	−0·020	0·005	0·000	
y_3	0·027	−0·019	−0·008	0·011	−0·019	0·000

Residuals ($\hat{\boldsymbol{\Sigma}} - \mathbf{S} - [(\bar{\mathbf{x}}, \bar{\mathbf{y}})' - \hat{\boldsymbol{\Lambda}}\hat{\boldsymbol{\theta}}][(\bar{\mathbf{x}}, \bar{\mathbf{y}})' - \hat{\boldsymbol{\Lambda}}\hat{\boldsymbol{\theta}}]'$)

x_1	3·764					
x_2	1·370	−0·413				
x_3	0·912	−0·258	−0·127			
y_1	−0·946	0·345	0·243	3·988		
y_2	0·777	−0·320	−0·157	1·338	−1·044	
y_3	0·868	−0·063	0·013	1·357	−0·698	−0·280

Table 3.—Derivatives and Residuals for the Model with ψ_{32} free

Derivatives

x_1	0·000					
x_2	−0·002	0·000				
x_3	0·003	0·000	0·000			
y_1	−0·046	0·024	0·025	0·000		
y_2	0·028	−0·014	−0·022	0·007	0·000	
y_3	0·037	−0·023	−0·012	0·013	−0·024	0·000

Residuals

x_1	0·433					
x_2	−0·015	−0·299				
x_3	0·043	−0·151	−0·072			
y_1	−1·062	0·062	0·064	3·824		
y_2	0·651	−0·575	−0·318	1·308	−1·034	
y_3	0·772	−0·262	−0·112	1·334	−0·722	−0·276

parameters again. At this step we find by Table 4 that the element ψ_{65} may not be equal to zero. A new estimation gives the result reported in Table 5. Now the fit of the model is acceptable, and it is seen that the original structure of the measurements has been retrieved.

The above procedure may be compared with that obtained by considering the residuals, which are given in Table 2 to Table 4. It will be seen that at no step would the correct covariance have been detected by choosing the largest off-diagonal residual. Further, it may seem irrational to relax only one restriction at each step, but an inspection of the information matrix shows that the estimates of the parameters in $\boldsymbol{\psi}$ are correlated. Thus, a procedure which relaxes more than one restriction at a time will be quite complex.

In applications with real data the χ^2 values for the different models may not indicate when to stop the model-fitting. In particular, when the number of observations is large the χ^2 values may well indicate that any conceptually

Table 4.—Derivatives and Residuals for the Model with ψ_{32} and ψ_{41} free

Derivatives						
x_1	0·000					
x_2	−0·002	0·000				
x_3	0·003	0·000	0·000			
y_1	0·000	0·001	−0·008	0·000		
y_2	0·001	0·000	−0·003	0·004	0·000	
y_3	0·004	−0·006	0·011	0·009	−0·016	0·000
Residuals						
x_1	0·719					
x_2	0·107	−0·263				
x_3	0·118	−0·123	−0·051			
y_1	1·666	0·152	0·120	4·201		
y_2	0·728	−0·492	−0·267	1·471	−0·948	
y_3	0·482	−0·197	−0·072	1·460	−0·604	−0·228

Table 5.—Parameter Estimates for the Final Model and their Standard Errors (within parentheses)

Factor loadings (Λ)		
x_1	1·000†	0·000†
x_2	0·809 (0·007)	0·000†
x_3	0·507 (0·004)	0·000†
y_1	0·000†	1·000†
y_2	0·000†	0·895 (0·004)
y_3	0·000†	0·695 (0·004)
Factor variance–covariance matrix (Φ)		
ξ	15·914 (0·794)	
η	11·711 (0·804)	25·640 (1·264)

Free parameters of ψ

$\hat{\sigma}_{\varepsilon_1}^2 = 3{\cdot}793\ (0{\cdot}341)$ $\quad\hat{\sigma}_{\varepsilon_3\varepsilon_2} = 0{\cdot}824\ (0{\cdot}129)$
$\hat{\sigma}_{\varepsilon_2}^2 = 2{\cdot}270\ (0{\cdot}220)$ $\quad\hat{\sigma}_{\varepsilon_1\delta_1} = 2{\cdot}395\ (0{\cdot}259)$
$\hat{\sigma}_{\varepsilon_3}^2 = 0{\cdot}968\ (0{\cdot}089)$ $\quad\hat{\sigma}_{\delta_3\delta_2} = 1{\cdot}361\ (0{\cdot}306)$
$\hat{\sigma}_{\delta_1}^2 = 6{\cdot}207\ (0{\cdot}536)$
$\hat{\sigma}_{\delta_2}^2 = 4{\cdot}497\ (0{\cdot}434)$
$\hat{\sigma}_{\delta_3}^2 = 2{\cdot}957\ (0{\cdot}271)$

Factor means

$\hat{E}(\xi) = 10{\cdot}002\ (0{\cdot}139)$
$\hat{E}(\eta) = 20{\cdot}007\ (0{\cdot}178)$

† Fixed parameters.

plausible model is non-acceptable (cf. Jöreskog, 1971, p. 421). In such a case examination of the difference in χ^2 values between two consecutive models in the procedure may help decide when elaboration of the model should stop. After the fourth step of the example, the derivatives suggest that ψ_{62} might be

different from zero (see Table 6). An estimation of this model results in a decrease in χ^2 by 0·506. Thus the hypothesis that $\psi_{62} = 0$ cannot be rejected. Since ψ_{62} is the parameter that gives the greatest decrease in χ^2, at least locally, it may be concluded that none of the other fixed ψ parameters is significantly different from zero.

Table 6.—Derivatives and Residuals for the Model with ψ_{32}, ψ_{41} and ψ_{65} free

Derivatives						
x_1	0·000					
x_2	−0·002	0·000				
x_3	0·003	0·000	0·000			
y_1	0·000	0·005	−0·003	0·000		
y_2	0·003	−0·003	−0·008	−0·001		
y_3	0·006	−0·010	0·006	0·001	0·000	0·000
Residuals						
x_1	0·808					
x_2	0·116	−0·361				
x_3	0·124	−0·186	−0·091			
y_1	1·186	−0·042	−0·002	1·213		
y_2	0·472	−0·704	−0·399	0·312	−0·469	
y_3	0·633	−0·362	−0·175	0·560	−0·124	0·090

3. An Application based on Longitudinal Growth Data

To exemplify the procedure described in the previous section, some data from the ETS longitudinal study of academic prediction and growth (cf. Anderson & Maier, 1963) are analysed. As part of that study, scores from the Sequential Test of Educational Progress (STEP) and School and College Ability Test (SCAT) were obtained. The data selected for the example concern 383 examinees for whom complete data are available, and three tests measuring verbal ability employed on two occasions. The following variables are identified:

x_1 = STEP Reading, Girls Academic, grade 9,
x_2 = STEP Writing, Girls Academic, grade 9,
x_3 = SCAT Verbal, Girls Academic, grade 9,
y_1 = STEP Reading, Girls Academic, grade 11,
y_2 = STEP Writing, Girls Academic, grade 11,
y_3 = SCAT Verbal, Girls Academic, grade 11.

In STEP Reading the testee is asked to read some written material and then to answer a number of multiple-choice items in connexion with each piece. Most of the items measure reading comprehension. In the items of STEP Writing the testee is usually asked to choose the correct evaluation of a given text with respect to grammatical or conceptual aspects. The SCAT Verbal test contains items of word completion and word synonyms. Thus, the information available about the measurements is rather complete, and it may

seem meaningless to use the procedure in this case, since it is merely constructed to be used in exploratory studies. On the other hand, the real data may be used as a test of whether the procedure produces a reasonable model or not.

As in Section 2 the first model analysed was a model with uncorrelated errors. By means of the χ^2 value for the overall fit of this model, as given in Table 7,

Table 7.—Goodness of Fit for the Different Models in the Application Example

Model	χ^2	d.f.	Probability level	χ^2 for test of the hypothesis $\psi_{ij} = 0$ (d.f. = 1)
Uncorrelated errors	74·219	12	0·000	
ψ_{63} free	27·393	11	0·004	46·826
ψ_{63}, ψ_{52} free	22·906	10	0·011	4·487

Table 8.—Derivatives for the Model with Uncorrelated Errors†

x_1	0·000					
x_2	−0·285	0·000				
x_3	0·201	0·092	0·000			
y_1	−0·126	−0·062	0·389	0·000		
y_2	−0·084	−0·167	0·179	−0·100	0·000	
y_3	0·466	0·291	−0·903	−0·010	0·115	0·000

† All values multiplied by 100.

it is concluded that the data do not fulfil the assumptions imposed. An inspection of the derivatives, shown in Table 8, suggests that the element ψ_{63} is not equal to zero. This means that there might be a correlation between the errors for the test SCAT Verbal across occasions. This is reasonable since the test SCAT Verbal, as compared with the other two tests, measures 'vocabulary' rather than 'a verbal ability'. Thus, when the influence of the verbal ability has been removed from the test scores at the two occasions, there might be covariation left due to the variation in vocabularly among the testees.

If the constraint $\psi_{63} = 0$ is relaxed a considerable decrease in χ^2 is obtained as indicated in Table 7. An inspection of the derivatives for this model, shown in Table 9, suggests that the element ψ_{52} is not equal to zero. This interpretation

Table 9.—Derivatives for the Model with ψ_{63} free†

x_1	0·000					
x_2	−0·122	0·000				
x_3	0·130	−0·005	0·000			
y_1	0·039	0·005	0·049	0·000		
y_2	0·042	−0·132	−0·121	0·043	0·000	
y_3	0·037	0·081	0·000	−0·101	0·058	0·000

† All values multiplied by 100.

of the derivatives is not as clear cut as in the previous case, and a release of the constraint $\psi_{52} = 0$ does not produce a significant decrease in χ^2 at the 1 per cent level, χ^2 with 1 degree of freedom being equal to 4·487. Further, the element ψ_{52} is estimated to be equal to 8·981 with a standard error equal to 4·363. Thus, in addition, the hypothesis that the separate parameter ψ_{52} is equal to zero cannot be rejected.

The estimates of the parameters in the model with ψ_{63} treated as a free parameter are listed in Table 10. It is seen that the covariance between the

Table 10.—Parameter Estimates for the Model with ψ_{63} treated as a Free Parameter, and Standard Errors of the Estimates (in parentheses)

Factor loadings (Λ)

STEP Reading$_9$	1·000†	0·000†
STEP Writing$_9$	0·995 (0·002)	0·000†
SCAT Verbal$_9$	0·968 (0·001)	0·000†
STEP Reading$_{11}$	0·000†	1·000†
STEP Writing$_{11}$	0·000†	0·990 (0·002)
SCAT Verbal$_{11}$	0·000†	0·955 (0·001)

Factor variance–covariance matrix (Φ) and factor means (θ)

	Variances and covariance		Means
Verbal ability$_9$	125·394 (9·914)		291·418 (0·643)
Verbal ability$_{11}$	131·979 (10·362)	151·403 (12·026)	303·057 (0·718)

Free parameters of Ψ

$\sigma_{\epsilon_1}^2 = 32{\cdot}753\ (3{\cdot}270)$ $\quad \sigma_{\epsilon_3\delta_3} = 14{\cdot}661\ (2{\cdot}521)$

$\sigma_{\epsilon_2}^2 = 73{\cdot}842\ (6{\cdot}045)$

$\sigma_{\epsilon_3}^2 = 25{\cdot}302\ (2{\cdot}745)$

$\sigma_{\delta_1}^2 = 45{\cdot}962\ (4{\cdot}384)$

$\sigma_{\delta_2}^2 = 65{\cdot}962\ (5{\cdot}658)$

$\sigma_{\delta_3}^2 = 33{\cdot}344\ (3{\cdot}470)$

† Fixed parameters.

errors for the test SCAT Verbal is considerable. In fact, the corresponding correlation coefficient is estimated as 0·505, although this is an 'attenuated' estimate of the 'vocabulary correlation', since the variances of the errors also contain measurement error variances.

The obtained model may be considered as a model with three common factors, the two ability factors plus a SCAT Verbal factor common to the two occasions. However, a model of this kind cannot be made identifiable in any reasonable way when the tests are known to contain measurement errors. All we can do is to include a term for the covariance between the errors to reduce the effect of this covariance on estimates of the parameters in the structural eqn. (2).

4. Some Statistical Properties of the Procedure

In this section the procedure exemplified in the previous sections is shown to be a special case of the Lagrangian multipliers test as described by Aitchison &

Silvey (1958) and Silvey (1958). The derivation given here is rather heuristic. For a rigorous derivation the reader is referred to the papers cited. It should be noted that the procedure is rather an *ad hoc* rule than a statistical procedure. Byron (1972) proposes a similar procedure for econometric systems and he recommends us not to use it too literally.

At each step of the procedure the introduction of a free parameter should be accompanied by an examination of its empirical interpretation. Thus the aim of the procedure is merely to generate hints in searching for the 'truth'. In any case, as already demonstrated in Section 2, it should be better than a procedure based on the estimated residuals.

In the most general case, the model considered in Sörbom (1974*b*) deals with the following equation for the observable variables, $\mathbf{x}$:

$$\mathbf{x}^{(g)} = \boldsymbol{\mu} + \boldsymbol{\Lambda}^{(g)}\boldsymbol{\xi}^{(g)} + \boldsymbol{\varepsilon}^{(g)} \quad (g = 1, 2, \ldots, m). \tag{7}$$

The superscript, g, refers to a division of the observations into m exclusive groups (cf. Sörbom, 1974*c*). The term $\boldsymbol{\xi}^{(g)}$ is a random vector of common factors assumed to be distributed as $N(\boldsymbol{\theta}^{(g)}, \boldsymbol{\Phi}^{(g)})$, $\boldsymbol{\varepsilon}^{(g)}$ is a random vector of unique factors, assumed to be uncorrelated with $\boldsymbol{\xi}^{(g)}$ and distributed as $N(\mathbf{0}, \boldsymbol{\Psi}^{(g)})$, $\boldsymbol{\mu}$ is a constant vector representing a common origin for the observable variables over the groups. $\boldsymbol{\Lambda}^{(g)}$ is a matrix of factor loadings.

The variance–covariance matrix for the observed variables in (7) is given by

$$\boldsymbol{\Sigma}^{(g)} = \boldsymbol{\Lambda}^{(g)}\boldsymbol{\Phi}^{(g)}\boldsymbol{\Lambda}^{(g)\prime} + \boldsymbol{\Psi}^{(g)}, \tag{8a}$$

and the vector of expected values by

$$E(\mathbf{x}^{(g)}) = \boldsymbol{\mu} + \boldsymbol{\Lambda}^{(g)}\boldsymbol{\theta}^{(g)}. \tag{8b}$$

The maximum-likelihood estimators are defined as those values of $\boldsymbol{\Lambda}^{(g)}$, $\boldsymbol{\Phi}^{(g)}$, $\boldsymbol{\Psi}^{(g)}$, $\boldsymbol{\theta}^{(g)}$ and $\boldsymbol{\mu}$ which maximize the likelihood function or, equivalently, the arguments to the function

$$F = \sum_{g=1}^{m} N^{(g)}[\log|\boldsymbol{\Sigma}^{(g)}| + \operatorname{tr}(\mathbf{T}^{(g)}\boldsymbol{\Sigma}^{(g)-1}) - \log|\mathbf{S}^{(g)}| - p] \tag{9}$$

at the minimum (cf. Sörbom, 1974*c*). $\mathbf{T}^{(g)}$ in (9) is the matrix

$$\mathbf{T}^{(g)} = \mathbf{S}^{(g)} + (\bar{\mathbf{x}}^{(g)} - \boldsymbol{\Lambda}^{(g)}\boldsymbol{\theta}^{(g)})(\bar{\mathbf{x}}^{(g)} - \boldsymbol{\Lambda}^{(g)}\boldsymbol{\theta}^{(g)})',$$

where $\mathbf{S}^{(g)}$ is the sample variance–covariance matrix for the gth group and $\bar{\mathbf{x}}^{(g)}$ is the vector of sample means. p is the number of observed variables and $N^{(g)}$ is the number of observations in the gth group.

The function F in (9) is a function of the elements of the matrices $\boldsymbol{\Lambda}^{(g)}$, $\boldsymbol{\Phi}^{(g)}$, $\boldsymbol{\Psi}^{(g)}$ and the vectors $\boldsymbol{\mu}$ and $\boldsymbol{\theta}^{(g)}$ for $g = 1, 2, \ldots, m$. Define $\boldsymbol{\pi}$ as the vector containing all these elements, in the following way

$$\begin{aligned}\boldsymbol{\pi} = (&\lambda_{11}^{(1)}, \lambda_{12}^{(1)}, \ldots, \lambda_{pk}^{(1)}, \phi_{11}^{(1)}, \phi_{21}^{(1)}, \ldots, \phi_{kk}^{(1)},\\ &\psi_{11}^{(1)}, \psi_{21}^{(1)}, \ldots, \psi_{pp}^{(1)}, \theta_{1}^{(1)}, \theta_{2}^{(1)}, \ldots, \theta_{k}^{(1)},\\ &\lambda_{11}^{(2)}, \lambda_{12}^{(2)}, \ldots, \ldots, \ldots, \theta_{k}^{(m)}, \mu_1, \mu_2, \ldots, \mu_p).\end{aligned}$$

The computer program described in Sörbom (1974*b*) minimizes F in (9) as a function of $\boldsymbol{\pi}$ under certain constraints on the elements of $\boldsymbol{\pi}$. The constraints are of two kinds, namely (i) fixed elements and (ii) equalities among elements. Thus the minimization procedure may be stated as follows. Find the minimum of $F(\boldsymbol{\pi})$ subject to the constraint $\mathbf{H}\boldsymbol{\pi} = \mathbf{s}$. This means that if the jth element of $\boldsymbol{\pi}$ belongs to category (i), then there is a 1 in the ith row and jth column of $\mathbf{H}$, i is the index running over the total number of constraints. All other elements of the ith row of $\mathbf{H}$ are zeros and s_i is the constant to which π_j is fixed. A constraint of the category (ii) gives rise to a row in $\mathbf{H}$ consisting of 1's in the positions corresponding to the elements of $\boldsymbol{\pi}$ involved in the constraint, except for one equal to -1 and zeros elsewhere. In this case s_j is equal to zero. If the Lagrangian multipliers $\mathbf{v}$ are introduced, the estimates of the parameters of the model are found as the solution of

$$\left.\begin{aligned}\partial F(\boldsymbol{\pi})/\partial\boldsymbol{\pi} + \mathbf{H}'\mathbf{v} &= \mathbf{0},\\ \mathbf{H}\boldsymbol{\pi} &= \mathbf{0}.\end{aligned}\right\} \tag{10}$$

As noted by Byron (1972), this implies that for a fixed element of $\boldsymbol{\pi}$ the derivative at the minimum is equal to minus the Lagrangian multiplier. Thus the procedure described in the previous sections is equivalent to selecting for relaxation that fixed element of $\boldsymbol{\pi}$ which has the greatest absolute value of its Lagrangian multiplier.

Byron (1972) proposes that we shall relax those fixed elements for which the ν_i's are significantly different from zero. To do this we have to find the standard errors for the vector $\mathbf{v}$. As shown by Silvey (1970), for example, this involves in principle the inversion of the information matrix for $\boldsymbol{\pi}$ and $\mathbf{v}$. This matrix is of order the total number of parameters in the model (7) minus the number of parameters constrained in order to make the model identified. Thus, the order of the matrix is so big that its storage may exhaust the memory of today's computers even for rather moderate-sized models. Further, if the null hypothesis, that is, the hypothesis that the data analysed are generated according to a specified model, is not accepted, then the estimates of $\mathbf{v}$ are biased, and Byron (1972, p. 755) concludes after a Monte Carlo study: 'it cannot be asserted that having rejected the null hypothesis, information can be gained from the rejected hypothesis about the true nature of the model which generated the sample'.

By the above, the *ad hoc* status of the procedure is apparent. It rather appeals to intuition in the sense that if a model is rejected we can find that fixed parameter which, when relaxed, gives a relatively large decrease in our measure of fit, the χ^2 value. It should be noted that there is no guarantee that we find the parameter that gives the greatest decrease, since, of course, we do not beforehand know the value of the parameter. A large change of one parameter with a small derivative could, at least theoretically, give rise to a greater decrease than a smaller change in another parameter with a large derivative. Thus the procedure shares the objection with the Lagrangian multiplier test, that as long as

we do not know the population parameters we cannot be certain to find the misspecified parameters. The procedure only gives hints, and as long as these hints give interpretable results we should accept them.

The research reported in this paper has been supported by the Swedish Council for Social Science Research under the project, Statistical Methods for Analysis of Longitudinal Data, project director K. G. Jöreskog.

References

AITCHISON, J. & SILVEY, S. D. (1958). Maximum-likelihood estimation of parameters subject to restraints. *Ann. math. Statist.* **29**, 813–828.

ANDERSON, S. B. & MAIER, M. H. (1963). 34,000 pupils and how they grow. *J. Teach. Educ.* **14**, 212–216.

ANDERSON, T. W. (1959). Some scaling models and estimation procedures in the latent class model. In U. Grenander (ed.), *Probability and Statistics.* New York: Wiley.

BYRON, R. P. (1972). Testing for misspecification in econometric systems using full information. *Int. econ. Rev.* **13**, 745–756.

CORBALLIS, M. C. (1973). A factor model for analysing change. *Br. J. math. statist. Psychol.* **26**, 90–97.

COSTNER, H. L. & SCHÖNBERG, R. (1973). Diagnosing indicator ills in multiple indicator models. In A. S. Goldberger & O. D. Duncan (eds.), *Structural Equation Models in the Social Sciences.* New York: Seminar Press.

JÖRESKOG, K. G. (1971). Simultaneous factor analysis in several populations. *Psychometrika* **36**, 409–426.

KENNY, D. A. (1973). Cross-lagged and synchronous common factors in panel data. In A. S. Goldberger & O. D. Duncan (eds.), *Structural Equation Models in the Social Sciences.* New York: Seminar Press.

SILVEY, S. D. (1958). The Lagrangian multiplier test. *Ann. math. Statist.* **30**, 389–407.

SILVEY, S. D. (1970). *Statistical Inference.* Harmondsworth: Penguin Books.

SÖRBOM, D. (1973). A statistical model for the measurement of change. Research Report 73–6, Statistics Department, Uppsala University.

SÖRBOM, D. (1974*a*). FASPSM: a computer program for factor analysis in several populations with structured means. (In preparation.)

SÖRBOM, D. (1974*b*). FASCOE: a computer program for factor analysis in several populations with structured means and correlated errors. (In preparation.)

SÖRBOM, D. (1974*c*). A general method for studying differences in factor means and factor structure between groups. *Br. J. math. statist. Psychol.* **27**, 229–239.

WILEY, D. E. & HORNIK, R. (1973). Measurement error and the analysis of panel data. Report no. 5, Studies of Educative Processes, University of Chicago.

Reprinted from: *British Journal of Mathematical and Statistical Psychology*, 1975, Vol. 28, pp. 138-151.

Part III
New Models for Group Comparisons